

STIC-Biotech/ChemLib

174960

From: Ramirez, Delia
Sent: Tuesday, December 20, 2005 4:43 PM
To: STIC-Biotech/ChemLib
Subject: 10/049,750

Hi,

I would like to request the following search:

1. a standard search of SEQ ID NO: 11, 13 and 14 in the protein databases (commercial & interference)

Thank you very much,

Delia M. Ramirez, Ph.D.
Patent Examiner
Recombinant Enzymes-Art Unit 1652
USPTO
400 Dulany Street, Remsen Bldg., 2D74, Mail room 2C70
Alexandria, VA 22314
(571) 272-0938
delia.ramirez@uspto.gov

RECEIVED
DEC 20 2005
STIC/CHEN. DIVISION
(STIC)

Searcher: _____
Searcher Phone: _____
Date Searcher Picked up: _____
Date completed: _____
Searcher Prep Time: _____
Online Time: _____

Type of Search
NA# _____ AA# _____
S/L: _____ Oligomer: _____
Encode/Transl: _____
Structure #: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other (Specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 23, 2005, 22:05:55 ; Search time 172.852 Seconds
(without alignments)
4682.260 Million cell updates/sec

Title: US-10-049-750-11

Perfect score: 1590

Sequence: 1 atggatgcgggtattgg.....atcttcgtgaacgaataa 921

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p model -DRV=xlp
-Q/cn2_1/USPTO_spool_p/US10049750/runat_23122005_113551_6285/app_query_faasta_1.1742
-DB=A Geneseq -QMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blom62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10049750 @CN 1.1 476 @runat_23122005_113551_6285 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_21.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1570	98.7	306	4 AAB67588	Aab67588 Amino aci
2	1519	95.5	306	4 AAB52865	Abd52865 Escherich
3	1207	75.9	311	7 ADF05492	Adf05492 Bacterial
4	870	54.7	298	8 ADS25796	Ads25796 Bacterial
5	870	54.7	298	8 ADS22364	Ads22364 Bacterial
6	870	54.7	298	8 ADS25248	Ads25248 Bacterial
7	870	54.7	298	8 ADS25978	Ads25978 Bacterial
8	831.5	52.3	321	4 AAU60867	Aau60867 Propionib
9	831.5	52.3	321	6 ABM57386	Abm57386 Propionib

10	797	50.1	304	8 ADN25077	Adn25077 Bacterial
11	550	34.6	308	4 AAU36261	Aau36261 Pseudomon
12	549	34.5	503	7 ABO83136	Abo83136 Pseudomon
13	546	34.3	322	4 AAB73533	Aab73533 Human tra
14	546	34.3	322	5 ABP55421	Abp55421 Human pro
15	546	34.3	322	8 ADO89874	Ado89874 Antagonis
16	546	34.3	322	8 ADR05197	Adr05197 Human rib
17	546	34.3	322	9 ADY17572	Ady17572 PRO polyp
18	542	34.1	322	4 AAM38771	Aam38771 Human pol
19	523	32.9	343	8 ADN22931	Adn22931 Bacterial
20	507	31.9	309	8 ADS23317	Ads23317 Bacterial
21	506	31.8	302	8 ADN24730	Adn24730 Bacterial
22	506	31.8	303	8 ADN21972	Adn21972 Bacterial
23	501	31.5	299	5 ABB90150	Abb90150 Human pol
24	499.5	31.4	300	8 ADN24942	Adn24942 Bacterial
25	495.5	31.2	303	8 ADS22183	Ads22183 Bacterial
26	495.5	31.2	300	8 ADS24911	Ads24911 Bacterial
27	491.5	30.9	320	7 ADF05698	Adf05698 Bacterial
28	488.5	30.7	293	8 ADN26498	Adn26498 Bacterial
29	476.5	30.0	304	4 AAU37424	Aau37424 Staphyloc
30	476.5	30.0	304	4 AAU37560	Aau37560 Staphyloc
31	474.5	29.8	304	6 ABM71981	Abm71981 Staphyloc
32	474.5	29.8	293	8 ADS44895	Ads44895 Bacterial
33	468.5	29.5	309	4 AAU38290	Aau38290 Salmonell
34	468.5	29.5	318	8 ADS44168	Ads44168 Bacterial
35	467	29.4	309	8 ADN17558	Adn17558 Bacterial
36	460	28.9	294	8 ADS28417	Ads28417 Bacterial
37	460	28.9	304	4 ABB68082	Abb68082 Drosophil
38	458	28.8	302	4 AAG82984	Aag82984 S. epider
39	457.5	28.8	293	8 ADS27571	Ads27571 Bacterial
40	454	28.6	303	8 ADS41653	Ads41653 Bacterial
41	452.5	28.5	330	7 ABO67526	Abo67526 Klebsiell
42	452	28.4	307	6 ABP19021	Abp19021 Pathogen
43	452	28.4	315	5 ABP39676	Abp39676 Staphyloc
44	452	28.4	315	8 ADS05785	Ads05785 Staphyloc
45	451	28.4	271	4 AAM40557	Aam40557 Human pol

ALIGNMENTS

RESULT 1

AAB67588

ID AAB67588 standard; protein; 306 AA.

XX

AC AAB67588;

XX

DT 29-MAY-2001 (first entry)

XX

DE Amino acid sequence of a deoxyribokinase enzyme.

XX

KW Deoxyribonucleoside; deoxyribose 1-phosphate; thymidine phosphorylase;
KW purine nucleoside phosphorylase; phosphopentose mutase;
KW phosphopentose aldolase; fructose 1,6-diphosphate aldolase;
KW deoxyribokinase; nucleoside 2-deoxyribosyltransferase.

XX Salmonella typhi.

OS

XX WO200114566-A2.

XX

PD 01-MAR-2001.

XX

PF 18-AUG-2000; 2000WO-EP008088.

XX

PR 20-AUG-1999; 99EP-00116425.

XX

PA (HOF) ROCHE DIAGNOSTICS GMBH.

PA (INSP) INST PASTEUR.

PA (PHAR-) PHARMA-WALDHOF GMBH & CO KG.

XX

PI Tischer W, Ihlenfeldt H, Barzu O, Sakamoto H, Pistotnik E;

PI Marliere P, Pochet S;

XX WPI; 2001-235026/24.

DR

DR N-PSDB; AAF55444.
 XX In vitro enzymatic synthesis of deoxyribonucleosides comprises reacting
 PT deoxyribose 1-phosphate and a nucleobase to form a deoxyribonucleoside
 PT and an inorganic phosphate.
 XX
 PS Disclosure; Page 59-61; 73pp; English.
 XX
 CC The present sequence represents a deoxyribokinase enzyme. This enzyme is
 CC involved in the biosynthesis of deoxyribonucleosides, and is used in the
 CC method of the invention. The specification describes a method for the in
 CC vitro enzymatic synthesis of deoxyribonucleosides. The method comprises
 CC reacting deoxyribose 1-phosphate and a nucleobase to form a
 CC deoxyribonucleoside and an inorganic phosphate. Enzymes which may be used
 CC in the method of the invention include thymidine phosphorylase, purine
 CC nucleoside phosphorylase, phosphopentose mutase, phosphopentose aldolase,
 CC fructose 1,6-diphosphate aldolase, deoxyribokinase, and nucleoside 2-
 CC deoxyribosyltransferase
 XX
 SQ Sequence 306 AA;

Alignment Scores:
 Pred. No.: 1.14e-168 Length: 306
 Score: 1570.00 Matches: 306
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 98.74% Indels: 0
 DB: 4 Gaps: 0

US-10-049-750-11 (1-921) x AAB67588 (1-306)

QY 1 ATGGATATCGCGTATTCGGCTCTACATGGTGGACCTTATCACCTACACCAACGATG 60
 DB 1 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrThrAsnGlnMet 20
 QY 61 CCCAAGAAGAGGGAACTCTGGAAGCGCGGGCTTTAAATCGGCTCGCGCGGAAAGGG 120
 DB 21 ProlysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyGlyGly 40
 QY 121 GCGAACACGCGCTGCGCGCGCTAAGCTCAATCAAAAGTATGTGATGTGACCAAGTG 180
 DB 41 AlaAsnGlnAlaValAlaAlaAlaLysLeuAsnSerLysValLeuMetLeuThrLysVal 60
 QY 181 GCGGAGATATTTTCCGACACACACCATTCGTATCTCGAATCTCGGGGGATCAATCG 240
 DB 61 GlyAspAspIlePheAlaAspAsnThrIleArgAsnLeuGluSerTrpGlyIleAsnThr 80
 QY 241 ACCTATGTAGAAAAGTACCGTGTACGACGCGCGGTAGCGCGGATTTTCGTCAACGCC 300
 DB 81 ThrTyrValGluLysValProCysThrSerSerGlyValAlaProIlePheValAsnAla 100
 QY 301 AACTCCAGCAACAGCATTCGATCATCAAGGCGCTAAACAAGTTCTCTCGCGGGAAGAT 360
 DB 101 AsnSerSerAsnSerIleLeuIleLysGlyAlaAsnLysPheLeuSerProGluAsp 120
 QY 361 ATCGATCGCGCGGCGAAGATTTAAAAATGCCAGCTTATTCGTCAACTGGGAAGTT 420
 DB 121 IleAspArgAlaAlaGluAspLeuLysCysGlnLeuIleValLeuGlnLeuGluVal 140
 QY 421 CAGCTTGAAACGGTTTATCAGCAATAGATTTGGCAAGAAACACCGGATTTGAAGTTGA 480
 DB 141 GlnLeuGluThrValTyrHisAlaIleGluPheGlyLysLysHisGlyIleGluValLeu 160
 QY 481 TTAACCCCTCGCGCACATTAACGGGAATTAGATATGCTCTTATGCTCTGTAATGCGATTTC 540
 DB 161 LeuAsnProAlaProAlaLeuArgGluLeuAspMetSerTyrAlaCysLysCysAspPhe 180
 QY 541 TTTGTACCTAATCAACCGCTGGAATATTAACCGGTATGCGATGATACCTATGAC 600
 DB 181 PheValProAsnGluThrGluLeuGluIleLeuThrGlyMecProValAspThrTyrAsp 200
 QY 601 CATATTTCGCGACGCGCATGTTTCGCTGGTAGATAAAGGGCTGAACAAATATTATTGTCACC 660

DB 201 HisIleArgAlaAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnIleIleValThr 220
 QY 661 ATGGCGGAGAAAGCGCGCTGTGGATGACGCTGACAGGAGTCCATGTTCCGCGCTTT 720
 DB 221 MetGlyGluLysGlyAlaLeuTrpMetThrArgAspGlnGluValHisValProAlaPhe 240
 QY 721 AGAGTGAACGCTGTTGTATACCGCGCGCGCGATGCTTTATCGGCTGTTTCGCGCAT 780
 DB 241 ArgValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheAlaHis 260
 QY 781 TACTACGTCCAGAGCGGGGATGTGAAGCGGCATGAAAAAGCCGCTCTCTTTCGCGCT 840
 DB 261 TyrTyrValGlnSerGlyAspValGluAlaMetLysLysAlaValLeuPheAlaA 280
 QY 841 TTCACGCTCACCGGAAAGGACCCCAATCTCTTATCAAGCATTGAGCAATTTAATGAG 900
 DB 281 PheSerValThrGlyLysGlyThrGlnSerSerTyrProSerIleGluGlnPheAsnGlu 300
 QY 901 TATCTTTCTGTTGAACGAA 918
 DB 301 TyrLeuSerLeuAsnGlu 306

RESULT 2
 ABB52865
 ID ABB52865 standard; protein; 306 AA.
 XX
 AC ABB52865;
 XX
 DT 11-FEB-2002 (first entry)
 XX
 DE Escherichia coli polypeptide SEQ ID NO 1139.
 XX
 KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
 KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
 KW systemic infection; non-diarrhoeal infection; septicemia;
 KW pyelonephritis; antibiotic resistance.
 XX
 OS Escherichia coli.
 XX
 PN WO200166572-A2.
 XX
 PD 13-SEP-2001.
 XX
 PF 12-MAR-2001; 2001WO-EP003445.
 XX
 PR 10-MAR-2000; 2000FR-00003145.
 PR 02-FEB-2001; 2001FR-00001449.
 XX
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX
 PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
 XX
 DR WPI; 2001-550253/61.
 XX
 PT A library of DNA fragments of Escherichia coli strains for the phylogenetic
 PT determination of a given strain comprises polynucleotides of nature B2/D+
 PT A-.

Example 6; Fig 6; 646pp; English.

The invention relates to a library of DNA fragments of Escherichia coli strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533) and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature B2/D+A-. The polynucleotides have potential antiinflammatory, antibacterial and immunosuppressive activity as part of pharmaceutical compositions used to treat, palliate or prevent extra-intestinal E. coli infections. The polypeptides are useful for determining the phylogenetic group of a given E. coli strain. These polypeptides can detect and treat an undesired development of E. coli, particularly an extra-intestinal infection that include systemic and non-diarrhoeal infections such as septicemia, pyelonephritis and meningitis this is particularly advantageous as bacterial resistance is increasing with the more frequent use of broad spectrum antibiotics


```
XX SQ Sequence 306 AA;
Alignment Scores:
Pred. No.: 7,13e-163 Length: 306
Score: 1519.00 Matches: 294
Percent Similarity: 98.69% Conservatives: 8
Best Local Similarity: 96.08% Mismatches: 4
Query Match: 95.53% Indels: 0
DB: 4 Gaps: 0

US-10-049-750-11 (1-921) x ABB52865 (1-306)
QY 1 ATGGATATCGCGTATTGGCTTAACATGGTGGACCTTATCACCTACCAACCAAGATG 60
Dd 1 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrThrAsnGlnMet 20
QY 61 CCCAAGAGAGGGAACCTCTGGAAGCGCGCGGTAAATCGGCTGCGCGGAAAGGG 120
Dd 21 ProLysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyLysGly 40
QY 121 GCGAACCGCGCGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTG 180
Dd 41 AlaAsnGlnAlaValAlaAlaAlaLysLeuAsnSerLysValLeuMetLeuThrLysVal 60
QY 181 GCGACGATATTTTGGCGACACACACATTTCGTAATCTCGAATCTCGGGGATCAATCG 240
Dd 61 GlyAspAspIlePheAlaAspAsnThrIleArgAsnLeuGluSerTrpGlyIleAsnThr 80
QY 241 AGTATGTAGAAAAGTACCTGTACAGAGCGCGGTAGCGCGATTTTCGTCAACGCC 300
Dd 81 ThrTyrValGluLysValProCysThrSerSerGlyValAlaProIlePheValAsnAla 100
QY 301 AACTCCAGCAACAGCATTCTGATCATCAAGGCGCTAAACAAGTTCTCTCGCGGAAGAT 360
Dd 101 AsnSerSerAsnSerIleLeuIleLysGlyAlaAsnLysPheLeuSerProGluAsp 120
QY 361 ATCGATCGCGCGCGGAAGATTTAAAAAATGCGAGCTTATTGTTCTGCAACTGGAAGTT 420
Dd 121 IleAspArgAlaAlaGluAspLeuLysLysCysLysLeuIleValLeuGlnLeuVal 140
QY 421 CAGCTTGAACGGTTTATCAGCAATAGAAATTGGCAAGAACACACGGGATTGAAGTTTA 480
Dd 141 GlnLeuGluThrValTyrHisAlaIleGluPheGlyLysLysIleGluValLeu 160
QY 481 TTAACCCCTGCGCGCAGCATTACGGGAATTAGATATCTTATGCTGTAATGCGGATTC 540
Dd 161 LeuAsnProAlaProAlaLeuArgGluLeuAspMetSerTyrAlaCysLysCysAspPhe 180
QY 541 TTGTGATCTTAATGAAACCGAGCTGGAATATTAAACCGGTATGCGAGTACCTATGAC 600
Dd 181 PheIleProAsnGluThrGluLeuGluIleLeuThrGlyMetSerValAspThrTyrAsp 200
QY 601 CATATTGCGCGACGCGCAGCTTCGCTGGTAGATAAAGGGCTGAACATATTATGTCACC 660
Dd 201 HisIleArgLeuAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnIleValThr 220
QY 661 ATGGCGGAGAAAGCGCGCTGTGGATGACCGGTGACCGAGAGTCCATGTCGGCGGTTT 720
Dd 221 MetSerGluLysGlyAlaLeuTrpMetThrArgAspGlnGluValHisValProAlaPhe 240
QY 721 AGAGTCAACGCTGTTGATACACGCGCGCGGCGATGCCCTTTATCGGCTGTTTCGCGCAT 780
Dd 241 LysValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheSerHis 260
QY 781 TACTACGTCCAGACGGGATGTTGGAAGCGGCATGAAAGCCGCTCTCTTTGCGCGT 840
Dd 261 TyrTy-ValGlnSerGlyAspValGluAlaAlaLeuLysLysAlaAlaLeuPheAlaAla 280
QY 841 TTCAGCGTACCGGGAAAGCCCAATCTCTTATCCAGCATTCAGCAATTTAATGAG 900
Dd 281 PheSerValThrGlyLysGlyThrGlnSerSerTyrProSerIleGluGlnPheAsnGlu 300
```

```
QY 901 TATCTTTCTGTTCAACGAA 918
Dd 301 PheLeuThrLeuAsnGlu 306

RESULT 3
ADF05492
ID ADF05492 standard; protein; 311 AA.
XX AC ADF05492;
XX AC ADF05492;
DT 12-FEB-2004 (first entry)
XX DE Bacterial polypeptide #1605.
XX KW Proteus mirabilis infection; bacterial infection; antibacterial;
XX KW immunostimulant.
XX OS Proteus mirabilis.
XX PN US6605709-B1.
XX PD 12-AUG-2003.
XX PF 05-APR-2000; 2000US-00543681.
XX PR 09-APR-1999; 99US-0128706P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Breton GL;
XX DR WPI; 2003-895291/82.
XX DR N-PSDB; ADF01320.
XX DX New Proteus mirabilis polypeptides and polynucleotides, useful as
PT reagents for diagnosis of bacterial disease, as components of
PT antibacterial vaccines, as targets for antibacterial drugs, or as
PT biocontrol agents for plants.
XX PS Disclosure; SEQ ID NO 5777; 870pp; English.
XX CC The invention relates to new Proteus mirabilis polypeptides and
CC polynucleotides. The invention also relates to antibodies against the
CC polypeptides, methods for producing the polypeptides, a method of
CC generating vaccines for immunising an individual against P. mirabilis, a
CC method for evaluating a compound for the ability to bind a P. mirabilis
CC polypeptide and a method for screening test compounds for anti-bacterial
CC activity. The polypeptides and polynucleotides are useful as molecular
CC targets for diagnosing, preventing and treating pathological conditions
CC resulting from bacterial infection, as reagents for diagnosis of
CC bacterial diseases, as components of antibacterial vaccines, as targets
CC for antibacterial drugs or as bio-control agents for plants. This
CC sequence represents a Proteus mirabilis polypeptide of the invention.
XX SQ Sequence 311 AA;

Alignment Scores:
Pred. No.: 2,05e-127 Length: 311
Score: 1207.00 Matches: 232
Percent Similarity: 87.79% Conservatives: 34
Best Local Similarity: 76.57% Mismatches: 37
Query Match: 75.91% Indels: 0
DB: 7 Gaps: 0

US-10-049-750-11 (1-921) x ADF05492 (1-311)
QY 1 ATGGATATCGCGTATTGGCTTAAACATGGTGGACCTTATCACCTACCAACCAAGATG 60
Dd 6 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrThrTyrIleAspArgMet 25
QY 61 CCCAAGAGAGGGAACCTCTGGAAGCGCGCGGTAAATCGGCTGCGCGGAAAGGG 120
Dd 26 ProLysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyLysGly 45
```

QY 121 GCGAACCGCGTGGCGCGCTAGCTCAATTCAAAAGATTGATGTTGACCAAGTG 180
Db 46 AlaAenGlnAlaValAlaAlaAlaLysLeuAenSerLysValMetMetLeuThrLysVal 65
QY 181 GCGCAGCATATTTTGGCCGACACACCATTCGTAATCTCGAATCTCTGGGGATCAATACG 240
Db 66 GlyAspAspIlePheAlaAspAenThrIleMetAenLeuGluSerTyrglyleAenThr 85
QY 241 AGCTATGTAGAAAAGTACCGGTGTACACGACGCGGTAGCGCGCATTTTCGTCAACGCC 300
Db 86 ArgTyrValGluLysValProGlyThrThrSerGlyValAlaProIlePheValThrSer 105
QY 301 AACTCCAGCAACAGCATCTCTGATCATCAAGAGCGCTAACAACTTCTCTCGCGGAAGAT 360
Db 106 GlnSerSerAsnSerIleLeuIleValIysGlyAlaAenGlnHisLeuSerAlaGluAsp 125
QY 361 ATCGATCGCGCGCGGAAGATTAAAAAAATCCAGCTTATTGTCTGCAACTGGAAGTT 420
Db 126 IleAspArgAlaAlaValSerLeuLysLysCysArgMetIleValLeuGlnLeuGluIle 145
QY 421 CAGCTTGAACCGTTTATCAGCAATAGAAATTTGGCAAGAAACACAGGGATTGAAAGTTTA 480
Db 146 ProLeuGluThrValTyrHisAlaIleGluPheGlyAsnLysHisIleProValLeu 165
QY 481 TTTAAACCTCGCGCCAGCATTCAGGGAATTAGATATGCTTATGCTCTGTAATGCGATTTC 540
Db 166 PheAenProAlaProAlaSerLysAlaLeuAspLeuThrIleAlaAlaGlnCysAspPhe 185
QY 541 TTTGTACCTTAATGAACCGAGCTGGAAATATTAAACCGGTATGCCAGTGGATACCTATGAC 600
Db 186 PheValProAsnGluThrGluLeuGluIleLeuThrGlyMetProIleAenThrLeuAsp 205
QY 601 CATATTCGCGCAGCGCATCGTTCGTGTGTAGATAAAGGCTGAAACAATATTATTGTACCC 660
Db 206 GluIleArgGluAlaAlaTyrPheLeuLeuGluLysGlyPheLysAsnIleIleValThr 225
QY 661 ATGCGCGAGAAAGCGCGCTGTGGATGACGCGTACCAGGAAGTCCATGTTCCGCGGTTT 720
Db 226 LeuGlyGluLysGlyAlaLeuIleThrValAenGlyGluIleGluLysTyrIleProAlaIle 245
QY 721 AGAGTGAACCGCTGTGATACACGCGCGCGCGATGCTTTATCGGCTGTTTTCGCGCAT 780
Db 246 GluValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheSerHis 265
QY 781 TACTACGTCAGCGGGGATGTGGAAGCGGCATGAAAAAAGCGCTCTCTTTCGCGCT 840
Db 266 TyrTyrValHisThrGlyAsnIleGluGluAlaLeuAsnLysAlaValMetPheSerGly 285
QY 841 TTCAGCGTCACCGGAAAGCGCCCAATCTCTTATCCAGCATTCGACCAATTTAATGAG 900
Db 286 LeuSerValThrGlyLysGlyThrGlnSerSerTyrProSerIleGluIlePheSerGlu 305
QY 901 TATCTTTTCG 909
Db 306 PheLeuSer 308

RESULT 4
ADS25796
ID ADS25796 standard; protein; 298 AA.
XX
AC ADS25796;
XX
DT 02-DEC-2004 (first entry)
XX

Bacterial polypeptide #14829.
Recombinant DNA construct; transformed plant; improved plant property;
cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
pathogen tolerance; pest tolerance; plant disease resistance;
cell cycle pathway modification; plant growth regulator;
homologous recombination; seed oil yield; protein yield; carbohydrate;
nitrogen; phosphorus; photosynthesis; lignin; galactomannan;

KW bacterial polypeptide.
XX Bacteria.
XX US2003233675-A1.
XX 18-DEC-2003.
XX 20-FEB-2003; 2003US-00369493.
XX 21-FEB-2002; 2002US-0360039P.
XX (CAOY/) CAO Y.
XX (HINK/) HINKLE G J.
XX (SLAT/) SLATER S C.
XX (CHEN/) CHEN X.
XX (GOLD/) GOLDMAN B S.
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
XX for expression of a polynucleotide encoding a polypeptide from a
XX microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 14829; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
XX promoter functional in a plant cell, where the promoter is positioned to
XX provide for expression of a polynucleotide encoding a polypeptide from a
XX microbial source. The invention also relates to a transformed plant
XX comprising the recombinant DNA construct and a method of producing a
XX transformed plant having an improved property. The plant is a crop plant
XX such as maize or soybean. The method of producing a transformed plant
XX having an improved property comprises transforming a plant with the
XX recombinant DNA construct and growing the transformed plant, where the
XX polynucleotide or polypeptide is useful for improving plant properties.
XX The recombinant DNA construct is useful for producing plants with
XX improved plant properties, e.g. improved cold, heat or drought tolerance,
XX tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
XX increased resistance to plant disease, better growth rate by modification
XX of the cell cycle pathway with plant growth regulators, increased rate of
XX homologous recombination, modified seed oil or protein yield and/or
XX content, improved yield by modification of carbohydrate, nitrogen or
XX phosphorus use and/or uptake, by modification of photosynthesis or by
XX providing improved plant growth and development under at least one stress
XX condition, improved lignin production or improved galactomannan
XX production. This sequence represents a bacterial polypeptide used in the
XX scope of the invention. Note: The sequence data for this patent did not
XX form part of the printed specification but was obtained in electronic
XX format from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 298 AA;

Alignment Scores:
Pred. No.: 3,998-89 Length: 298
Score: 870.00 Matches: 167
Percent Similarity: 76.61% Conservative: 59
Best Local Similarity: 56.61% Mismatches: 69
Query Match: 54.72% Indels: 0
DB: 8 Gaps: 0

US-10-049-750-11 (1-921) x ADS25796 (1-298)

QY 7 ATCGCGGTTATTGGCTTAACATGGTGACCTTATCACCTACCAACAGATGCCAAA 66
Db 4 IleGlyValValGlySerAsnMetValAspLeuIleThrTyrValAspArgMetProGly 23
QY 67 GAAGGGGAAACTCTGGAACCGCGCGGTTTAAATCGGCTCGCGGAAAGGGCGGAC 126
Db 24 ProGlyGluThrLeuGluAlaProThrPheGluMetGlyCysGlyGlyLysGlyAlaAsn 43


```
Db 64 AspValPheAlaAspAsnThrIleArgAsnLeuAlaSerPheGlyValAspThrArgHis 83
QY 247 GTAGAAAAGTACCGGTACCGAGCGCGGTAGCGCGATTTTGTCTCAACGCCACTCC 306
Db 84 ValValIleValSerGlyLysSerGlyValAlaProIlePheValGluGlnSerGly 103
QY 307 AGCAACAGCATTTCTGATCATCAAGCGCTAAACAGTTTCTCTCGCGCGAAGATATCGAT 366
Db 104 GluAsnSerIleLeuIleValIleValIleGlyAlaAsnAlaAspLeuLeuProValGluValAsp 123
QY 367 CGCGCGCGGAGATTTAAAAAATGCCAGCTTATTGTTCTGCAATGGAAGTTGAGTTT 426
Db 124 LysAlaAlaAspLeuLysGluCysGlyLeuIleLeuMetGlnMetGluValProVal 143
QY 427 GAAACGGTTTATCAGCATAGATTTGGCAGAAACACGCGATTTGAAGTGTATTAAAC 486
Db 144 GluThrValIleHisThrIleGluPheAlaAlaGlnAsnGlyIleGluThrIleLeuAsn 163
QY 487 CTGCGCCAGCATTAACGGAAATTAGATATGTTCTATGCTGTAAATGCGATTTCTTTGTA 546
Db 164 ProAlaProAlaAlaAsnLeuAspProGluIleArgGlnValThrPheLeuVal 183
QY 547 CCTAATGAAACCGAGTGGAAATATTACCGGTATGCCAGTGGATACCTATGACCATATT 606
Db 184 ProAsnGluSerGluLeuAlaLeuLeuSerGlyLeuProThrAspThrAspGluAspIle 203
QY 607 CGCGACGCGCGCTGCGTGGTAGATAAGGCTGAAACAATATTATTCACCATGGCG 666
Db 204 ValArgAlaAlaArgSerLeuIleAlaArgGlyIleArgThrValIleValThrLeuGly 223
QY 667 GAGAAAGCGCGTGTGGATGACGCTGACAGGAAGTCCATGTTCCGCGTTTGAAGTG 726
Db 224 AlaArgGlyAlaArgMetIleThrSerAspGluIleValAsnIleGluProValIleVal 243
QY 727 AACGCTGTTGTACACCGCGCGCGCGATGCTTTATCGGTGTTTTCGCGATTACTAC 786
Db 244 ThrProArgAspThrThrGlyAlaGlyAspAlaPheIleGlySerPheAlaArgPheTyr 263
QY 787 GTTCAGAGCGGGATGTGAAGCGCGCATGAAAGCGCTCTCTTTCGCGCTTTCAGC 846
Db 264 AlaGluThrArgGluValValSerSerLeuLysLysAlaSerLeuTyrAlaAlaHisSer 283
QY 847 GTCACCGGGAAGGACCACTCTCTTATCCAGCATTGAGCAA 891
Db 284 IleThrArgProGlyThrGlnLysAlaIleAlaSerIleAspGlu 298

RESULT 6
AD525248
ID ADS25248 standard; protein; 298 AA.
XX AC ADS25248,
XX DT
XX DE
XX DE Bacterial polypeptide #14281.
XX KW Recombinant DNA construct; transformed plant; improved plant property;
XX cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
XX pathogen tolerance; pest tolerance; plant disease resistance;
XX cell cycle pathway modification; plant growth regulator;
XX homologous recombination; seed oil yield; protein yield; carbohydrate;
XX nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
XX OS Bacteria.
XX PN US2003233675-A1.
XX XX
XX PD 18-DEC-2003.
XX PF 20-FEB-2003; 2003US-00369493.
XX XX
```

```
PR 21-FEB-2002; 2002US-0360039P.
XX (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
DR
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX Claim 1; SEQ ID NO 14281; 122pp; English.
PS
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
XX format from USPTO at seqdata.uspto.gov/sequence.html.
SQ Sequence 298 AA;
Alignment Scores:
Pred. No.: 3,998-89 Length: 298
Score: 870.00 Matches: 167
Percent Similarity: 76.61% Conservative: 59
Best Local Similarity: 56.61% Mismatches: 69
Query Match: 54.72% Indels: 0
DB: Gaps: 0
US-10-049-750-11 (1-921) x ADS25248 (1-298)
QY 7 ATCGCGGTTATTGGCTCTAAACATGGTGACCTTATCACCTACCAACAGATGCCCAA 66
Db 4 IleGlyValIleGlySerAsnMetValAspLeuIleThrTyrValAspArgMetProGly 23
QY 67 GAAGGGGAAACTCTGGAAGCGCGCGTTTAAATTCGGTCTGGCGGAAAGGGGGAAC 126
Db 24 ProGlyGluThrLeuGluAlaProThrPheGluMetGlyCysGlyGlyLysGlyAlaAsn 43
QY 127 CAGGCGGTGGCGCGCTAAGCTCAATTCAAAAGTATTGATCTTGACCAAGTGGCGGAC 186
Db 44 GlnAlaValAlaAlaAlaArgLeuGlyAlaGluValMetMetValThrArgValGlyAsp 63
QY 187 GATATTTTTCGCGACAAACACCATTCGTATCTCGAATCTCGAGGATCAATACGAGTAT 246
Db 64 AspValPheAlaAspAsnThrIleArgAsnLeuAlaSerPheGlyValAspThrArgHis 83
QY 247 GTAGAAAAGTACCGGTACCGAGCGCGGTAGCGCGATTTTCTCTCAACGCCACTCC 306
XX
```

Db 84 ValValLysValSerGlyLysSerSerGlyValAlaProIlePheValGluGlnSerGly 103
QY 307 AGCAACAGCATTTCTGATCATCAAGCGCTAAACAGTTTCTCTCGCGGGAAGATATCGAT 366
Db 104 GluAenSerIleLeuIleValLysGlyAlaAenAlaAspLeuLeuProValGluValAsp 123
QY 367 CGCGCGCGGAGAGATTAAAAATCCAGCTTATTGTTCTGCACTGGAAGTTGAGTT 426
Db 124 LysAlaAlaAspLeuLysGlyCysGlyLeuIleLeuMetGlnMetGluValProVal 143
QY 427 GAAACGGTTTATCACCAATAGATTGGCAAGAACCGGATTCAGGTATTATTAAAC 486
Db 144 GluThrValTyHisThrIleGluPheAlaAlaGlnAsnGlyIleGluThrIleLeuAen 163
QY 487 CTGCGCCAGCATTCAGGAATTAGATATGCTTATGCTCTGTAATGCGATTCTTTGTA 546
Db 164 ProAlaProAlaAlaAenLeuAspProGluArgIleArgGlnValThrPheLeuVal 183
QY 547 CTTATGAAACGAGCTGGAATATTAAACCGGTATGCCAGTGTGATCCTATGACCATATT 606
Db 184 ProAsnGluSerGluLeuAlaLeuLeuSerGlyLeuProThrAspThrAspGluAspIle 203
QY 607 CGCGCAGCGGCGTTCGCTGCTAGATTAAGGCTGAACAATATTATTCTCACCAGGCG 666
Db 204 ValArgAlaAlaArgSerIleAlaArgGlyIleArgThrValIleValThrLeuGly 223
QY 667 GAGAAGGCGCGTGTGGATGACGCGTACAGGAAGTCCATGTTCCGCGTTTGAAGTG 726
Db 224 AlaArgGlyAlaArgMetIleThrSerAspGluIleValAsnIleGluProValLysVal 243
QY 727 AACGCTGTTGATACAGCGCGCGGCGATGCTTTATCGCTGTTTTCGCGCATTTACTAC 786
Db 244 ThrProArgAspThrThrGlyAlaGlyAspAlaPheIleGlySerPheAlaArgPheTyr 263
QY 787 GTCCAGCGCGGATGTGGAACCGCCATGCAAAAGCGCTCTCTTTCGCGCTTTCAGC 846
Db 264 AlaGluThrArgGluValValSerSerLeuLysLysAlaSerLeuTyrAlaAlaHisSer 283
QY 847 GTCCAGCGGAAAGGCAACCAATCTCTTATCCAAGCATTTGAGCAA 891
Db 284 IleThrArgProGlyThrGlnLysAlaTyrAlaSerIleAspGlu 298

RESULT 7

ADS25978

ID ADS25978 standard; protein; 298 AA.

XX

AC ADS25978;

XX

DT 02-DEC-2004 (first entry)

XX Bacterial polypeptide #15011.

XX Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.

XX Bacteria.

XX US2003233675-A1.

XX

XX

XX

XX

XX 20-FEB-2003; 2003US-00369493.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PA

PA

XX

PI

XX

XX

DR

XX

XX

PT

PT

PT

XX

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

(CHEN/) CHEN X.

(GOLD/) GOLDMAN B S.

Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

WPI; 2004-061375/06.

New recombinant DNA construct comprising a promoter positioned to provide

for expression of a polynucleotide encoding a polypeptide from a

microbial source, useful for producing plants with improved properties.

Claim 1; SEQ ID NO 15011; 122pp; English.

The invention relates to a recombinant DNA construct comprising a

promoter functional in a plant cell, where the promoter is positioned to

provide for expression of a polynucleotide encoding a polypeptide from a

microbial source. The invention also relates to a transformed plant

comprising the recombinant DNA construct and a method of producing a

transformed plant having an improved property. The plant is a crop plant

such as maize or soybean. The method of producing a transformed plant

having an improved property comprises transforming a plant with the

recombinant DNA construct and growing the transformed plant, where the

polynucleotide or polypeptide is useful for improving plant properties.

The recombinant DNA construct is useful for producing plants with

improved plant properties, e.g. improved cold, heat or drought tolerance,

tolerance to herbicides, extreme osmotic conditions, pathogens or pests,

increased resistance to plant disease, better growth rate by modification

of the cell cycle pathway with plant growth regulators, increased rate of

homologous recombination, modified seed oil or protein yield and/or

content, improved yield by modification of carbohydrate, nitrogen or

phosphorus use and/or uptake, by modification of photosynthesis or by

providing improved plant growth and development under at least one stress

condition, improved lignin production or improved galactomannan

production. This sequence represents a bacterial polypeptide used in the

scope of the invention. Note: The sequence data for this patent did not

form part of the printed specification but was obtained in electronic

format from USPTO at seqdata.uspto.gov/sequence.html.

Sequence 298 AA;

Alignment Scores:

Pred. No.: 3 99e-89 Length: 298

Score: 870.00 Matches: 167

Percent Similarity: 76.61% Conservative: 59

Best Local Similarity: 56.61% Mismatches: 69

Query Match: 54.72% Indels: 0

DB: Gaps: 0

US-10-049-750-11 (1-921) x ADS25978 (1-298)

QY 7 ATCGCGGTTATTGGCTTAACATGGTGGACCTTATCACCAACAGATGCCAA 66

Db 4 IleGlyValValGlySerAsnMetValAspLeuIleThrTyValAspArgMetProGly 23

QY 67 GAAGGGGAAACTCTGAGACGGCGCGGTTTAAATCGCTGGCGGGAAGGGGGAAC 126

Db 24 ProGlyGluThrLeuGluAlaProThrPheGluMetGlyCysGlyGlyLysGlyAlaAsn 43

QY 127 CAGGCGGCGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGGAC 186

Db 44 GlnAlaValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 63

QY 187 GATATTTTTCGCGACAAACACCATTCGTATCTCGAATCTCGGGGATCAATACGAGTAT 246

Db 64 AspValPheAlaAspAsnThrIleArgAsnLeuAlaSerPheGlyValAspThrArgHis 83

QY 247 GTAGAAAAGTACCGGTGTACAGCGCGCGGTAGCGCCGATTTCTGTCACGCCAATCC 306

Db 84 ValValLysValSerGlyLysSerGlyValAlaProIlePheValGluGlnSerGly 103

QY 307 AGCAACAGCATTTCTGATCATCAAGCGCTAAACAGTTTCTCTCGCGGGAAGATATCGAT 366

Db 104 GluAenSerIleLeuIleValLysGlyAlaAenAlaAspLeuLeuProValGluValAsp 123

Qy	661	ATGGCGGAGAAAGCGCGCTGTGGATGACGCCTGACCAGGAAGTCAVTTTCGGCGTTT	720
Db	237	LeuGlySerArGlyAlaMetTrpAlaHisAlaGluGlyArgLysIleileLysAlaPro	256
Qy	721	AGAGTGACCGCTGTGATACAGCGCGCGCGGATGCTTTATCGGCTGTTTCGGCAT	780
Db	257	ValValGlnAlaValAspThrThrGlyAlaGlyAspAlaPheIleGlyCysPheAlatys	276
Qy	781	TACTACGTCCAGAGCGCGGATGTGGAAGCGGCATGAAAAAAGCCGTCCTCTTTTGC CGCT	840
Db	277	GluTrpValAspHisGlyAspValIleAlaIleArgAlaGlyAsnArgTyzrAlaAla	296
Qy	841	TTACGCGTCACCGGAAAAGGCCACCAATCCTCTTAT	876
Db	297	AspSerValThrArgHisGlyThrGlnSerSerTyr	308
RESULT 10			
ID	ADN25077		
XX	ADN25077 standard; protein; 304 AA.		
AC	ADN25077;		
XX			
DT	02-DEC-2004 (first entry)		
XX	Bacterial polypeptide #7730.		
DE			
KW	Recombinant DNA construct; transformed plant; improved plant property;		
KW	cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;		
KW	pathogen tolerance; pest tolerance; plant disease resistance;		
KW	cell cycle pathway modification; plant growth regulator;		
KW	homologous recombination; seed oil yield; protein yield; carbohydrate;		
KW	nitrogen; phosphorus; photosynthesis; lignin; galactomannan;		
KW	bacterial polypeptide.		
OS	Bacteria.		
XX			
PN	US2003233675-A1.		
XX			
PD	18-DEC-2003.		
XX			
PF	20-FEB-2003; 2003US-00369493.		
XX			
PR	21-FEB-2002; 2002US-0360039P.		
XX			
PA	(CAOY/) CAO Y.		
PA	(HINK/) HINKLE G J.		
PA	(SLAT/) SLATER S C.		
PA	(CHEN/) CHEN X.		
PA	(GOLD/) GOLDMAN B S.		
PI	Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;		
DR	WPI; 2004-061375/06.		
XX			
PT	New recombinant DNA construct comprising a promoter positioned to provide		
PT	for expression of a polynucleotide encoding a polypeptide from a		
PT	microbial source, useful for producing plants with improved properties.		
XX	Claim 1; SEQ ID NO 7730; 122pp; English.		
XX			
CC	The invention relates to a recombinant DNA construct comprising a		
CC	promoter functional in a plant cell, where the promoter is positioned to		
CC	provide for expression of a polynucleotide encoding a polypeptide from a		
CC	microbial source. The invention also relates to a transformed plant		
CC	comprising the recombinant DNA construct and a method of producing a		
CC	transformed plant having an improved property. The plant is a crop plant		
CC	such as maize or soybean. The method of producing a transformed plant		
CC	having an improved property comprises transforming a plant with the		
CC	recombinant DNA construct and growing the transformed plant, where the		
CC	polynucleotide or polypeptide is useful for improving plant properties.		
CC	The recombinant DNA construct is useful for producing plants with		
CC	improved plant properties, e.g. improved cold, heat or drought tolerance.		

QY 727 AACGCTGTTGATACACAGCGCGCGGATGCTTTATCGGCTGTTTCGGCGATTACTAC 786
Db 255 ThrProValAspThrThrGlyAlaGlyAspAlaPheIleGlyAlaPheAlaHisPheLeu 274
QY 787 GTCCAGAGCGGGATGTGGAAGCGCGCATGAAAAAGCCGCTCTTTGCGCGCTTTCAGC 846
Db 275 AlaAlaThrGlyGluValGluGlyAlaLeuAlaHisAlaAlaArgTyrAlaAlaHisSer 294
QY 847 GTCACCGGGAAGACCCCAATCTCTTAT 876
Db 295 ValThrGlyArgGlyThrGlnLysSerTyr 304.

RESULT 11

AAU36261

ID AAU36261 standard; protein; 308 AA.

XX AC

XX AAU36261;

XX DT 14-FEB-2002 (first entry)

XX XX

DE Pseudomonas aeruginosa cellular proliferation protein #251.

XX Antisense; prokaryotic cellular proliferation protein; antibiotic;

XX antibacterial; drug design.

XX Pseudomonas aeruginosa.

XX WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US009180.

XX PR 21-MAR-2000; 2000US-0191078P.

XX PR 23-MAY-2000; 2000US-0206848P.

XX PR 26-MAY-2000; 2000US-0207272P.

XX PR 23-OCT-2000; 2000US-0242578P.

XX PR 27-NOV-2000; 2000US-0253625P.

XX PR 22-DEC-2000; 2000US-0257931P.

XX PR 16-FEB-2001; 2001US-0269308P.

XX XX

PA (ELIT-) ELITRA PHARM INC.

XX XX

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;

PI Yamamoto RT, Xu HH;

PI WPI; 2001-611495/70.

DR N-PSDB; AAS54120.

XX XX

PT New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids.

XX XX

PS Example 3; SEQ ID NO 11854; 511pp; English.

XX XX

CC The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the genes,

CC themselves and the encoded proteins. The prokaryotes used are Escherichia

CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,

CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also

CC useful for the identification of potential new targets for antibiotic

CC development. The antisense nucleic acids can also be used to identify

CC proteins used in proliferation, to express these proteins, and to obtain

CC antibodies capable of binding to the expressed proteins. The proteins can

CC be used to screen compounds in rational drug discovery programmes. The

CC antisense nucleic acid sequence is also useful to screen for homologous

CC nucleic acids which are required for cell proliferation in a wide variety

CC of organisms. The present sequence represents an essential prokaryotic

CC cellular proliferation protein. Note: The sequence data for this patent

CC did not form part of the printed specification, but was obtained in

CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX XX

SQ Sequence 308 AA;

Alignment Scores:

Pred. No.: 9.41e-53 Length: 308

Score: 550.00 Matches: 126

Percent Similarity: 59.33% Conservative: 52

Best Local Similarity: 42.00% Mismatches: 122

Query Match: 34.59% Indels: 0

DB: 4 Gaps: 0

US-10-049-750-11 (1-921) x AAU36261 (1-308)

QY 7 ATCGCGGTTATTGGCTCTAACATGTTGGACCTTATCACCTACCAACACAGATGCCAAA 66

Db 5 ValLeuValValGlySerLeuAenMetAspLeuValValArgAlaProArgLeuProArg 24

QY 67 GAAGGGAAACTCTGGAAGCGCGCTTTAAATCGGCTGGCGGCGGAAAGGGGGAAC 126

Db 25 GlyGlyGluThrLeuAlaGlyGlnSerPheIleThrValProGlyGlyGlyAlaAen 44

QY 127 CAGCGCGTGGCGCGCTTAAGCTCAATTCAAAGTATTGATGTTGACCAAGTGGCGGAC 186

Db 45 GlnAlaValAlaAlaAlaArgLeuGlyAlaGluValAlaMetIleGlyCysLeuGlyAsp 64

QY 187 GATATTTTGGCGCAACACCATTCGTAATCTCGAATCTGGGGGATCAATACGACGTAT 246

Db 65 AspAlaTyrGlyAspGlnLeuTyrArgAlaLeuGlnAlaGluGlyIleAspCysGlnGly 84

QY 247 GTAGAAAAAGTACCGGTACACAGCGCGCTAGCGCCGATTTTCGTCAACGCCAACTCC 306

Db 85 ValGluArgValAlaGlyGlnSerSerGlyValAlaLeuIleValValAspAspSerSer 104

QY 307 AGCAACAGCATTTCTGATCATCAAGCGCTAAACAGTTTCTCTCGCGGGAAGATATCGAT 366

Db 105 GlnAenAlaIleValIleValAlaGlyGlyAsnGlyHisLeuSerProAlaValLeuAla 124

QY 367 CGCGCGCGGAGAGATTAAAAAATCCAGCTTATTGTTCTGCAACTGGAGTTCAGCTT 426

Db 125 ArgHisGluHisLeuLeuGluGlnAlaGlnValAlaValCysGlnLeuGluSerProLeu 144

QY 427 GAAACGGTTTATCACGCAATAGAAATTTGGCAAGAAACACGGGATTCGAAGTGTATTAAAC 486

Db 145 GluThrValGlyHisValLeuArgArgAlaHisAlaLeuGlyLysThrValIleLeuAen 164

QY 487 CTGCGCCAGCATTTACGGAATTTAGATATGCTTTATGCTCTGTAATATTCGATTTCTTGTGA 546

Db 165 ProAlaProAlaThrArgAspValProAlaGluTrpLeuProLeuValAspTyrLeuVal 184

QY 547 CTTAATGAAACCGAGCTGGAATATTAACGGGTATGCCAGTGGATACCTATGACCATATT 606

Db 185 ProAenGluThrGluSerGluLeuLeuCysArgLeuProValAspSerLeuGluSerAla 204

QY 607 CGCGCAGCGGCGGCTTCGCTGTAGATAAAGGCTGAACAATATTATTGTCACATGGGC 666

Db 205 GlyArgAlaAlaGluArgLeuArgGluMetGlyAlaGlyArgValIleValThrLeuGly 224

QY 667 GAGAAGGCGCGCTGTGGATGACCGCTGACAGGAGTCCATGTTCTCCGCGGTTTACAGTG 726

Db 225 AlaGlnGlyAlaLeuLeuValGlyGlyArgValGluHisPheProValAlaArgVal 244

QY 727 AACGCTGTTGATACAGCGCGCGGCGGATGCTTTATCGGCTGTTTCCGGCATTTACTAC 786

Db 245 LysAlaLeuAspThrThrAlaAlaGlyAspThrPheValGlyGlyPheAlaAlaLeu 264

QY 787 GTCCAGAGCGGAGTGTGGAAGCGCCATGAAAAAGCCGCTCTCTTTCGCGCTTTTCAGC 846

Db 265 AlaArgGlyLeuAspGluAlaAlaIleArgPheGlyGlnAlaAlaAlaIleSer 284

QY 847 GTCACCGGGAAGACCCCAATCTCTTATCCAAGCATTTGAGCAATTAATAGTATCTTT 906

Db 285 ValThrArgLeuGlyAlaGlnThrSerIleProSerArgGluGluValGluArgAlaLeu 304

KW gene therapy; drug screening.

XX Homo sapiens.

XX WO200132888-A2.

XX 10-MAY-2001.

XX 02-NOV-2000; 2000WO-US030485.

XX 04-NOV-1999; 99US-0163595P.

XX (INCY-) INCYTE GENOMICS INC.

XX Tang YT, Yue H, Hillman JL, Lal P, Bandman O, Patterson C;

PI Shih LL, Azimzai Y, Lu DAM, Baughn MR;

XX WPI; 2001-328796/34.

DR N-PSDB; AAH23840.

XX Human transferase polypeptides and polynucleotides useful for diagnosis, prevention and treatment of cell proliferative and immune system disorders and for identifying agonists and antagonists.

PS Claim 1; Page 133-134; 157pp; English.

XX Sequences AAB73494-AAB73535 represent novel human transferase proteins HTFS-1 to HTFS-42, and sequences AAH23801-AAH23842 represent cDNAs encoding them. The proteins play important roles in the regulation of cellular signalling and proliferation. The HTFS proteins are useful for screening compounds for their effectiveness as agonists or antagonists of transferase activity, or for compounds that specifically bind to an HTFS protein or which modulates the activity of an HTFS protein.

XX Pharmaceutical compositions comprising an HTFS protein. HTFS agonist or antagonist, or genetic construct encoding an HTFS protein are useful for treating a disease or condition associated with decreased or increased expression of functional HTFS. Disorders which may be treated using such compositions include cell proliferative disorders and immune disorders. For example, diseases which may be treated include atherosclerosis, hepatitis, psoriasis, cancers (including breast, bladder, bone marrow, brain and uterus cancer), inflammation, AIDS, Addison's disease, allergies, asthma, anaemia, cirrhosis, Crohn's disease, atopic dermatitis, diabetes mellitus, multiple sclerosis, rheumatoid arthritis, pancreatitis, systemic lupus erythematosus, thrombocytopenia, and ulcerative colitis. They may also be used to treat complications of cancer, haemodialysis, extracorporeal circulation, trauma and haematopoietic cancer, including lymphoma, leukaemia and myeloma.

XX Polynucleotides encoding HTFS proteins are useful for creating transgenic animals to model human diseases, for diagnostic purposes and to generate hybridisation probes useful in mapping the naturally occurring genomic sequences. HTFS, and its catalytic or immunogenic fragments are useful for screening libraries of compounds in a variety of drug screening techniques. Antibodies which specifically bind HTFS may be used for the diagnosis of disorders associated with the expression of HTFS, or in assays to monitor patients being treated with HTFS or agonists, antagonists or inhibitors of HTFS. The present sequence represents an HTFS protein of the invention

XX Sequence 322 AA;

Alignment Scores:

Pred. No.:	2,74e-52	Length:	322
Score:	546.00	Matches:	117
Percent Similarity:	58.98%	Conservative:	57
Best Local Similarity:	39.66%	Mismatches:	117
Query Match:	34.34%	Indels:	4
DB:		Gaps:	2

US-10-049-750-11 (1-921) x AAB73533 (1-322)

QY 7 ATCGCGGTATTGGCTTAACTATGGAGCCTTATCATCAACCAACAGATGCCAAA 66

Db 18 ValValValValGlySerCysMetThrAspLeuValSerLeuThrSerArgLeuProlys 37

QY	67	GAAGGGGAAACTCTGGAAGCGCGCGGCTTTAAATAATCGGTGGCGGAAAGGGGCGAAC	126
Db	38	ThrGlyGluThrIleHisGlyHisLysPhePheIleGlyPheGlyGlyGlyAlaAsn	57
QY	127	CAGGCGGTGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGGGAC	186
Db	58	GlnCysValGlnAlaAlaArgLeuGlyAlaMetThrSerMetValCysLysValGlyLys	77
QY	187	GATATTTTGGCGCAACACCACTTCGAATCTCGAATCTGGGGGATCAATACGCTAT	246
Db	78	AspSerPheGlyAsnAspTyrIleGluAsnLeuLysGlnAsnAspIleSerThrGluPhe	97
QY	247	GTAGAAAAGTACCGGTGTACACGACGCGGTAGCCGCTTTCCTCAACGCCAACTCC	306
Db	98	ThrTyrGlnThrLysAspAlaAlaThrGlyThrAlaSerIleIleValAsnAsnGluGly	117
QY	307	AGCAACAGCATTCGTGATCATCAAGGCGCTTAACAGTTTCTCTCGCGGAGATATCGAT	366
Db	118	GlnAsnIleIleValIleValAlaGlyAlaAsnLeuLeuAsnThrGluAspLeuArg	137
QY	367	CGCGCGCGGAAAGATTAAAAAAATGCCAGCTTATTGTTCTGCAACTGGAAGTTAGCTT	426
Db	138	AlaAlaAlaAsnValIleSerArgAlaLysValMetValCysGlnLeuGluIleThrPro	157
QY	427	GAACCGGTTTATCACGCAATAGAAATTTGGCAAGAAACACGCGGATTGAAGTGTATTAAAC	486
Db	158	AlaThrSerLeuGluAlaLeuThrMetAlaAlaArgSerGlyValLysThrLeuPheAsn	177
QY	487	CTGCGCGCAGCATTCAGGGAATTAGATATGCTTATGCTGTAAATGCGATTCTTTGTA	546
Db	178	ProAlaProAlaIleAlaAlaAspLeuAspProGlnPheTyrThrLeuSerAspValPheCys	197
QY	547	CCTAATGAACCGAGCTGCAATATTAACCGGTATGCCAGTGGATACCTATGACCATATT	606
Db	198	CysAsnGluSerGluAlaGluLeuThrGlyLeuThrValGlySerAlaAlaAspAla	217
QY	607	CGCGCAGCGCGCGCTTCGCTGGTAGATAAAGGCGTGAAACAATATTATTGTCCACATGGGC	666
Db	218	GlyGluAlaAlaLeuValLeuLysArgGlyCysGlnValValIleIleThrLeuGly	237
QY	667	GAGAAGGC-----GCGCTGGTAGACGCGTGACCGAGAGTTCATGTCGGCGTTT	720
Db	238	AlaGluGlyCysValValLeuSerGlnThrGluProGluProLysHisIleProThrGlu	257
QY	721	ACAGTGAACGCTGTTGATACACGCGCGCGCGATGCTTATCGGCTGTTTCGCGCAT	780
Db	258	LysValLysAlaAlaValAspThrThrGlyAlaGlyAspSerPheValGlyAlaLeuAlaPhe	277
QY	781	TAC-----TACGTCCAGACGCGGATGTGGAAGCGCGCATGAAAGGCGCTCTCTTT	834
Db	278	TyrLeuAlaTyrTyrProAsnLeuSerLeuGluAspMetLeuAsnArgSerAsnPheIle	297
QY	835	GCGCGTTTCAGCGTCACCGGAAAGGACCCCAATCTCTTATCCA 879	
Db	298	AlaAlaValSerValGlnAlaAlaGlyThrGlnSerSerTyrPro 312	

RESULT 14

ABP55421

ID ABP55421 standard; protein; 322 AA.

XX

AC ABP55421;

XX

DT 04-FEB-2003 (first entry)

XX

DE Human protein 35.42 SEQ ID NO:2.

XX

KW Human; protein 35.42; pfkB protein family; tumour; infection; dementia;

KW embryonic development disorder.

XX

OS Homo sapiens.

XX

PN CN1342697-A.

XX 03-APR-2002.
XX PD
XX PF 12-SEP-2000; 2000CN-00125166.
XX PR 12-SEP-2000; 2000CN-00125166.
XX PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX PI Mao Y, Xie Y;
XX WPI; 2002-529774/57.
XX N-PSDB; ABQ83903.
XX Human protein 35.42 containing pfkB protein family characteristic
PT sequence fragment and encoding polynucleotide, useful for treating
PT embryonic development disorder, tumor, infection and dementia.
XX
XX Claim 1; Page 26-27 (Disclosure); 33pp; Chinese.
XX
XX The present sequence represents human protein 35.42 containing pfkB
CC protein family characteristic sequence fragments (I). Also described is a
CC method for preparing (I) using DNA recombination technology. (I) can be
CC used for treating embryonic development disorders, tumors, infections
CC and dementia
XX
XX Sequence 322 AA;
SQ

Alignment Scores:
Pred. No.: 2,74e-52 Length: 322
Score: 546.00 Matches: 117
Percent Similarity: 58.98% Conservative: 57
Best Local Similarity: 39.66% Mismatches: 117
Query Match: 34.34% Indels: 4
DB: Gaps: 2

US-10-049-750-11 (1-921) x ABP5421 (1-322)

QY 7 ATCGCGGTTATGGCTCTAACATGCTGGAGCTTATCACCCTACACCAACAGATGCCCAA 66
DB 18 ValValValValGlySerCysMetThrAspLeuValSerLeuThrSerArgLeuProLys 37
QY 67 GAAGGGAACTCTGAAGCGCGGCTTTAAATCGCTGCGCGGAAAGGGGCGAAC 126
DB 38 ThrGlyGluThrIleHisGlyHisLysPhePheIleGlyPheGlyLysGlyAlaAsn 57
QY 127 CAGCGCGTGGCGCGCGCTCAAGCTCAATCAAAAGTATTGATGTTGACCAAGTGGCGGAC 186
DB 58 GlnCysValGlnAlaAlaArgLeuGlyAlaMetThrSerMetValCysLysValGlyLys 77
QY 187 GATATTTTGGCGACCAACACCATTCGTAATCTCGAATCTCGGGGATCAATACGACGTAT 246
DB 78 AspSerPheGlyAsnAspTyrIleGluAsnLeuLysGlnAsnAspIleSerThrGluPhe 97
QY 247 GTAGAAAAGTACCGTGTACACGCGGCTAGCGCGATTTTCTGTCACAGCCAACTCC 306
DB 98 ThrTyrGlnAlaLysAspAlaAlaThrGlyThrAlaSerIleIleValAsnAsnGluGly 117
QY 307 AGCAACAGACGATCTGATCATCAAGCGCTAAACAGTTTCTCTCGCGGGAAGATATCGAT 366
DB 118 GlnAsnIleIleValIleValAlaGlyAlaAsnLeuLeuAsnThrGluAspLeuArg 137
QY 367 CGCGCGGCGAAGATTTAAAAAATGCCAGCTTATTGTTCTGCACTGGAAGTTACGCTT 426
DB 138 AlaAlaAlaAsnValIleSerArgAlaLysValMetValCysGlnLeuGluIleThrPro 157
QY 427 GAAACCGTTATCACCAATAGATTGGCAAGAAACACCGGATTCGAAGTGTATTAAAC 486
DB 158 AlaThrSerLeuGluAlaLeuThrMetAlaArgArgSerGlyValLysThrLeuPheAsn 177
QY 487 CTGCGCGCCAGCATTAACGGGAATTAGATATGCTCTATGCTCCCTGTAATGCGGATTTCTTGA 546
DB 178 ProAlaProAlaIleAlaAspLeuAspProGlnPheTyrThrLeuSerAspValPheCys 197

QY 547 CCTAATGAACCGAGCTGGAAATATTAAACCGGTATGCCAGTGGATACCTATGACCATATT 606
DB 198 CysAsnGluSerGluAlaGluIleLeuThrGlyLeuThrValGlySerAlaAlaAspAla 217
QY 607 CGCGCAGCGGACGCTTCCTGCTAGATAAAGCGGTGAACAATATTATTGTCACCATGGGC 666
DB 218 GlyGluAlaAlaLeuValLeuLysArgGlyCysGlnValValIleIleThrLeuGly 237
QY 667 GAGAAAGGC-----GCGCTGTGGATGACGCGTGACCGAGAGTCCATGTTCCGCGGTTT 720
DB 238 AlaGluGlyCysValValLeuSerGlnThrGluProGluProLysHisIleProThrGlu 257
QY 721 AGAGTGAACGCTGTTGATACACGCGCGCGCGATGCTTATCGGCTGTTTCGCGCAT 780
DB 258 LysValLysAlaValAspThrThrGlyAlaGlyAspSerPheValGlyAlaLeuAlaPhe 277
QY 781 TAC-----TACGTCCAGAGCGGGATGTGGAAGCGGCATGAAAAAGCCGCTCTCTTT 834
DB 278 TyrLeuAlaTyrTyrProAsnLeuSerLeuGluAspMetLeuAsnArgSerAsnPheIle 297
QY 835 GCGGCTTTCAGCGTCCACCGGGAAGCACCACCAATCTCTTATCCA 879
DB 298 AlaAlaValSerValGlnAlaAlaGlyThrGlnSerSerTyrPro 312

RESULT 15
ADQ89874
ID ADQ89874 standard; protein; 322 AA.
XX AC ADQ89874;
XX DT 21-OCT-2004 (first entry)
XX DE Antagonist of cell cycle progression polypeptide #152.
XX KW Cytostatic; cancer; cell division cycle; mitosis; meiosis;
XX OS Homo sapiens.
XX PN WO2004063362-A2.
XX PD 29-JUL-2004.
XX PF 31-DEC-2003; 2003WO-GB005635.
XX PR 10-JAN-2003; 2003US-0439123P.
XX PR 06-MAY-2003; 2003US-0468402P.
XX PA (CYCL-) CYCLACEL LTD.
XX PI Glover D, Bell G, Frenz L, Midgley C;
XX DR WPI; 2004-544089/52.
XX DR N-PSDB; ADQ89873.
XX PT New cell cycle progression genes and proteins for modulating cell cycle
PT progression in cells, for preventing, treating or diagnosing cell
PT proliferative diseases (e.g. cancer) or for identifying modulators of
PT mitosis or meiosis.
XX PS Claim 2; SEQ ID NO 304; 461pp; English.
XX CC The present invention relates to a polynucleotide for preventing,
CC treating or diagnosing a disease in an individual. The composition or the
CC polypeptide, polynucleotide or RNA precursor, or antibody is useful for
CC diagnosing, preventing or treating diseases (e.g. cell proliferative
CC diseases such as cancer) in an individual. These may also be used for
CC identifying substances capable of binding to or modulating the function
CC of the polypeptide, capable of affecting the function of the
CC corresponding gene, or capable of inhibiting the cell division cycle or
CC cell cycle progression, preferably mitosis and/or meiosis. The present
CC sequence represents an antagonist of cell cycle progression protein

CC sequence.

XX Sequence 322 AA;

Search completed: December 23, 2005, 22:29:50
Job time : 186.852 secs

Alignment Scores:

Pred. No.	2,74e-52	Length:	322
Score:	546.00	Matches:	117
Percent Similarity:	58.98%	Conservative:	57
Best Local Similarity:	39.66%	Mismatches:	117
Query Match:	34.34%	Indels:	4
DB:	8	Gaps:	2

US-10-049-750-11 (1-921) x ADO89874 (1-322)

Qy	7	ATCGCGGTATTGGCTCTAAACATGCTGGACCTTATCATCCTACACCAACAGATGCCCAA	56
Db	18	ValValValValGlySerCysMetThrAspLeuValSerLeuThrSerArgLeuProLys	37
Qy	67	GAACGGGAAACTCTCGAAGCCGCGCTTTAAATTCGGCTGCGCGGAGAAAGGGCGCAAC	126
Db	38	ThrGlyGluThrIleHisGlyHisLysPhePheIleGlyPheGlyGlyLysGlyAlaAsn	57
Qy	127	CAGGCGGTGGCGCGCTAAAGCTCAATTCAAAGATTGATGTATCCACCAAGCTGGCGCAC	186
Db	58	GlnCysValGlnAlaAlaArgLeuGlyAlaMetThrSerMetValCysLysValGlyLys	77
Qy	187	GATATTTTGGCGCAACACCATTCGTATCTGAATCTCGGAATCTGGGGATCAATACGAGTAT	246
Db	78	AspSerPheGlyAsnAspTyrIleGluAsnLeuLysGlnAsnAspIleSerThrGluPhe	97
Qy	247	GTAGAAAAGTACCGTGTACACGACAGCGGTAGCGCGGATTTTCGTCAACGCCAATCC	306
Db	98	ThrTyrGlnThrLysAspAlaAlaThrGlyThrAlaSerIleValAsnAsnGluGly	117
Qy	307	AGCAACACGATTCGTATCATCAACACGCGCTAACAAAGTTTCTCGCGGAAGATATCGAT	366
Db	118	GlnAsnIleIleValIleValAlaGlyAlaAsnLeuLeuLeuLeuLeuThrGluAspLeuArg	137
Qy	367	CGCGCGCGGAAGATTAAAAAATCCGACGCTTATTGTTCTCAACTGGAAGTTCAGCTT	426
Db	138	AlaAlaAlaAsnValIleSerArgAlaLysValMetValCysGlnLeuGluIleThrPro	157
Qy	427	GAACAGGTTTATACGCAATAGAAATTGGCAAGAAACACGGGATTGAAGTGTTATTAAAC	486
Db	158	AlaThrSerLeuGluAlaLeuThrMetAlaArgArgSerGlyValLysThrLeuPheAsn	177
Qy	487	CCTCGCCAGCATTACGGGAATTAGATATGCTTATGCTGTAAGTGCATTCCTTTGTA	546
Db	178	ProAlaProAlaIleAlaAspLeuAspProGlnPheTyrThrLeuSerAspValPheCys	197
Qy	547	CCTAATGAACCGAGCTGGAATATTAAACGGGTATGCCAGTGCATCTACCATATT	606
Db	198	CysAsnGluSerGluAlaGluIleLeuThrGlyLeuThrValGlySerAlaAlaAspAla	217
Qy	607	CGCGCAGCGCACGTTTCGCTGTAGATAAAGGGCTGCAACAATATTATTGTACCAATGGCG	666
Db	218	GlyGluAlaAlaLeuValLeuLeuLysArgGlyCysGlnValValIleIleThrLeuGly	237
Qy	667	GAGAAAGC-----GCGCTGTGGATGACGCGCTGCACGAGGAAGTCCATGTTCCGGCGTTT	720
Db	238	AlaGluGlyCysValValLeuSerGlnThrGluProGluProLysHisIleProThrGlu	257
Qy	721	AGAGTGAACGCTGTGTATACACGCGCGCGCGGATGCGCTTTATCGGCTGTTTCGCGCAT	780
Db	258	LysValLysAlaValAspThrThrGlyAlaGlyAspSerPheValGlyAlaLeuAlaPhe	277
Qy	781	TAC-----TACGTCACGAGCGGGATGTGGAAAGCGGCCCATAAAAAACCCGCTCTTT	834
Db	278	TyrLeuAlaTyrTyrProAsnLeuSerLeuGluAspMetLeuAsnArgSerAsnPheIle	297
Qy	835	GCCGCTTTACGCGTCACCGGGAAAGCAACCCCAATCTCTTTATCCA	879
Db	298	AlaAlaValSerValGlnAlaAlaGlyThrGlnSerSerTyrPro	312

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 23, 2005, 22:17:41 ; Search time 37.391 Seconds
(without alignments)
4072.865 Million cell updates/sec

Title: US-10-049-750-11
Perfect score: 1590
Sequence: 1 atggatcgcgggtattgg.....atcttccttgacgaataa 921

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 572060 seqs, 82675679 residues
Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlp
-MODEL=frame+n2p.model
-O=/cgn2_1/USPTO spool_p/US10049750/runat_23122005_113553_6317/app query fasta_1.1742
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=n2p.ra -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10049750@cgn 1 1 101 @runat 23122005_113553_6317 -NCPU=6 -ICPU=3
-NO KWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/1aa/5 COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/6 COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/H COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/PTCUT COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/RE COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1207	75.9	311	2	US-09-543-681A-5777
2	549	34.5	503	2	US-09-252-991A-31882
3	491.5	30.9	320	2	US-09-543-681A-5983
4	458	28.8	302	2	US-09-710-279-3062
5	452.5	28.5	330	2	US-09-489-039A-14043
6	452	28.4	315	2	US-09-134-001C-4521
7	447.5	28.1	311	2	US-09-543-681A-7498
8	420.5	26.4	406	2	US-09-543-681A-8124
9	387.5	24.4	342	2	US-09-489-039A-9668
10	377.5	23.7	328	2	US-09-489-039A-7290
11	376.5	23.7	529	2	US-09-134-000C-4524
12	310.5	19.5	451	2	US-09-543-681A-4841

13	250	15.7	170	2	US-09-107-532A-3946	Sequence 3946, Ap
14	234	14.7	319	2	US-09-710-279-786	Sequence 786, Ap
15	233	14.7	321	2	US-09-134-001C-3900	Sequence 3900, Ap
16	220.5	13.9	269	2	US-09-134-000C-5214	Sequence 5214, Ap
17	217.5	13.7	331	2	US-09-543-681A-7111	Sequence 7111, Ap
18	215	13.5	328	2	US-08-826-611-6	Sequence 6, Appli
19	215	13.5	362	2	US-09-328-352-7264	Sequence 7264, Ap
20	214.5	13.5	313	2	US-09-489-039A-12895	Sequence 12895, A
21	211.5	13.3	347	2	US-08-826-611-2	Sequence 2, Appli
22	198.5	12.5	324	2	US-09-489-039A-13558	Sequence 13558, A
23	194	12.2	308	2	US-08-913-816C-17	Sequence 17, Appl
24	191	12.0	372	2	US-08-918-249-2	Sequence 2, Appli
25	191	12.0	372	2	US-08-918-249-4	Sequence 4, Appli
26	191	12.0	372	2	US-09-345-603-2	Sequence 2, Appli
27	191	12.0	372	2	US-09-345-603-4	Sequence 4, Appli
28	189.5	11.9	330	2	US-08-936-165A-384	Sequence 384, App
29	189.5	11.9	331	2	US-09-902-540-10221	Sequence 10221, A
30	183.5	11.5	303	2	US-08-913-816C-2	Sequence 2, Appli
31	182.5	11.5	356	2	US-09-489-039A-12757	Sequence 12757, A
32	160	10.1	115	2	US-09-599-360B-81	Sequence 81, Appl
33	160	10.1	115	2	US-09-621-976-5617	Sequence 5617, Ap
34	160	10.1	341	2	US-09-489-039A-8625	Sequence 8625, Ap
35	156.5	9.8	314	2	US-09-489-039A-10377	Sequence 10377, A
36	154	9.7	324	2	US-09-328-352-7458	Sequence 7458, Ap
37	152.5	9.6	303	1	US-08-961-539-2	Sequence 2, Appli
38	152.5	9.6	303	2	US-09-185-826-2	Sequence 2, Appli
39	148.5	9.3	491	2	US-09-543-681A-6634	Sequence 6634, Ap
40	148	9.3	334	2	US-09-107-532A-6336	Sequence 6336, Ap
41	146.5	9.2	312	2	US-09-489-039A-8270	Sequence 8270, Ap
42	146.5	9.2	477	2	US-09-492-709A-325	Sequence 325, App
43	145.5	9.2	303	2	US-09-583-110-5072	Sequence 5072, Ap
44	145.5	9.2	311	2	US-09-107-433-4423	Sequence 4423, Ap
45	144	9.1	526	2	US-09-252-991A-32984	Sequence 32984, A

ALIGNMENTS

RESULT 1

US-09-543-681A-5777
; Sequence 5777, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543, 681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5777
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5777

Alignment Scores:
Pred. No.: 1.18e-130 Length: 311
Score: 1207.00 Matches: 232
Percent Similarity: 87.79% Conservative: 34
Best Local Similarity: 76.57% Mismatches: 37
Query Match: 75.91% Indels: 0
DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x US-09-543-681A-5777 (1-311)

QY	1	ATFGATATCGCGTATTGGCTTAACATGTTGGACCTTATCACCTTACCAACACGATG	60
Db	6	MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrTyrlleAspArgMet	25
QY	61	CCCAAGAGGGGAACCTCTGGAAGCGCGGCTTTAAATCGCTCGCGCGAAGGG	120

```
Db      26  ProLysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyGlyLysGly 45
Qy      121  GGAACACCGCGCTGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTG 180
Db      46  AlaAsnGlnAlaValAlaAlaAlaLysLeuAsnSerLysValMetMetLeuThrLysVal 65
Qy      181  GCGCAGCATATTTTGGCCACACACCATTCGTAATCTCGAATCTCTGGGGGATCAATACG 240
Db      66  GlyAspAspIlePheAlaAspThrIleMetAsnLeuGluSerTyrglyIleAsnThr 85
Qy      241  ACGTATGTAGAAAAAGTACCGTGTACACAGCGCGGTAGCCCGATTTTCGTCAACGCC 300
Db      86  ArgTyrValGluLysValProGlyThrThrSerGlyValAlaProIlePheValThrSer 105
Qy      301  AACTCAGCAACAGCATTTCTGATCATCAAGGGCTTAACAAGTTTCTCTCGCGGAGAT 360
Db      106  GlnSerSerAsnSerIleLeuIleValLysGlyAlaAsnGlnHisLeuSerAlaGluAsp 125
Qy      361  ATCGATCGCGCGCGAAGATTAAAAAAATGCCAGCTTATTGTTCTGCAACTGGAAGTT 420
Db      126  IleAspArgAlaAlaValSerLeuLysLysCysArgMetIleValLeuGlnLeuGluIle 145
Qy      421  CAGCTTGAAACGGTTTATCACCAATAGAAATTGGCAAGAAACACCGGGATTGAAGTTTA 480
Db      146  ProLeuGluThrValTyrHisAlaIleGluPheGlyAsnLysHisIleProValLeu 165
Qy      481  TTAACCCCTGCGCAGCATTCAGGNAATTAGATATGTCTTATGCCGTGTAATAACGCAATTC 540
Db      166  PheAsnProAlaProAlaSerLysAlaLeuAspLeuThrIleAlaAlaGlnCysAspPhe 185
Qy      541  TTTGTACCTTAATCAAAACCGAGCTGGAATATTAAACCGGTATGCCAGTGGATACCTATGAC 600
Db      186  PheValProAsnGluThrGluLeuGluIleLeuThrGlyMetProIleAsnThrLeuAsp 205
Qy      601  CATATTGCGCGCAGCGCAGCTTCGCTGGTAGATAAGGCTGAACAATATTATGTCACC 660
Db      206  GluIleArgGluAlaAlaTyrPheLeuLeuGluLysGlyPheLysAsnIleIleValThr 225
Qy      661  ATGGGGGAGAAAGCGCGCTGGATGACGCGTGACCGAGTCCATGCTTCGGGGTTT 720
Db      226  LeuGlyGluLysGlyAlaLeuTrpValAsnGlyGluIleGluLysTyrlleProAlaIle 245
Qy      721  AGAGTGAACGCTGTTCATACACGCGCGCGGCGATGCTTATCGCTGTTTCGGGCAT 780
Db      246  GluValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheSerHis 265
Qy      781  TACTACGTCAGAGCGGGATGTGGAAGCGCGCATGAAAAAGCCGCTCTCTTCGCCGT 840
Db      266  TyrTyrValHisThrGlyAsnIleGluGluAlaLeuAsnLysAlaValMetPheSerGly 285
Qy      841  TTCAGCGTCACCGGGAAGCACCAGCCCAATCTCTTATCCAGCAATTGAGCAATTAATGAG 900
Db      286  LeuSerValThrGlyLysGlyThrGlnSerSerTyrlleGluLeuGluPheSerGlu 305
Qy      901  TATCTTTTCG 909
Db      306  PheLeuSer 308
```

RESULT 2

```
US-09-252-991A-31882
; Sequence 31882, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
```

```
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31882
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31882

Alignment Scores:
Pred. No.: 2,73e-54 Length: 503
Score: 549.00 Matches: 126
Percent Similarity: 59.14% Conservative: 52
Best Local Similarity: 41.84% Mismatches: 123
Query Match: 34.53% Indels: 0
DB: 2 Gaps: 0
```

US-10-049-750-11 (1-921) x US-09-252-991A-31882 (1-503)

```
Qy      7  ATFCGCGTTATTTGGCTTAACATGTTGGACCTTATCACCCTACACCAACAGATGCCAAA 66
Db      200  ValLeuValValGlySerLeuAsnMetAspLeuValValArgAlaProArgLeuProArg 219
Qy      67  GAAGGGGAACTCTCGAAGCGCGCGTAAATCGGCTGGCGGAAAGGGGCGAC 126
Db      220  GlyGlyGluThrLeuAlaGlyGlnSerPheThrThrIleProGlyGlyLysGlyAlaAsn 239
Qy      127  CAGGCGGTGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGGAC 186
Db      240  GlnAlaValAlaAlaAlaArgLeuGlyAlaGlyValAlaMetIleGlyCysLeuGlyAsp 259
Qy      187  GATATTTTGGCGACACACCATTCGTATCTCGAATCTCGGAGTCAATACGAGTAT 246
Db      260  AspAlaTyrGlyAspGlnLeuTyrArgAlaLeuGlnAlaGluGlyIleAspCysGlnGly 279
Qy      247  GTAGAAAAGTACCGTGTACACGACGCGGTAGCGCGATTTTCGTCAACGCCCACTCC 306
Db      280  ValGluArgValAlaGlyGluSerSerGlyValAlaLeuIleValValAspAspSerSer 299
Qy      307  AGCAACAGCATTCATCATCAAGCGCTAAACAGTTTCTCTCGCGGAGATATTCGAT 366
Db      300  GlnAsnAlaIleValIleValAlaGlyGlyAsnGlyHisLeuSerProAlaValLeuAla 319
Qy      367  CGCGCGCGGAGAGATTTAAAAAATGCCAGCTTATTGTTCTGCAACTGGAAGTTCAGCTT 426
Db      320  ArgHisGluHisLeuLeuGluGlnAlaGlnValValValCysGlnLeuGluSerProLeu 339
Qy      427  GAAACGGTTTATCACCAATAGAAATTTGCAAGAAACACGCGGATTGAAGTGTATTAAAC 486
Db      340  GluThrValGlyHisValLeuArgAlaHisAlaLeuGlyLysThrValIleLeuAsn 359
Qy      487  CTGCGCCAGCATTTACGGGAATTAGATATGCTTATGCTCTGTAATGCGATTTCTTTGTA 546
Db      360  ProAlaProAlaThrArgAspValProAlaAspTrpLeuProLeuValAspTyrLeuVal 379
Qy      547  CCTAATGAAACAGAGCTGGAATATTAAACCGGTATGCCAGTGCATACCTATGACCATATT 606
Db      380  ProAsnGluThrGluSerGluLeuLeuCysArgLeuProValAspSerLeuGluSerAla 399
Qy      607  CGCGCAGCGGCACGTTTCGCTGAGTAAAGGCTGAACAATATTATTCCTCACCATGGGC 666
Db      400  GlyArgAlaAlaGluArgLeuArgGluMetGlyAlaGlyArgValIleValThrLeuGly 419
Qy      667  GAGAAAGCGCGCTGTGGATGACGCGTGACAGGAAGTCCATGTTCTCGCGGTTTACAGTG 726
Db      420  AlaGlnGlyAlaLeuLeuValGlyGluGlyArgValGluHisPheProValAlaArgVal 439
Qy      727  AACGCTGTTGATACCAAGCGCGCGCGATGCGCTTTATCGGCTGTTTCGCGCATTTACTAC 786
Db      440  LysAlaLeuAspThrThrAlaAlaGlyAspThrPheValGlyGlyPheAlaAlaLeu 459
Qy      787  GTCCAGAGCGGGGATGTGGAAGCCCATGAAAGCCGCTCTCTCTTCGCGCTTTTCAGC 846
Db      460  AlaArgGlyLeuAspGluAlaAlaIleArgPheGlyGlyGlnAlaAlaAlaIleSer 479
```


QY 847 GTCCACGGGAAAGGACCCCAATCTCTTATCCAGCATTTGAGCAATTTAATGATATCTT 906
Db 480 ValThrArgLeuGlyAlaGlnThrSerIleProSerArgGluGluValGluArgAlaLeu 499
QY 907 TCG 909
Db 500 Ala 500

RESULT 3

US-09-543-681A-5983
; Sequence 5983, Application US/09543681A
; Patent No. 6605709

GENERAL INFORMATION:
; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A

; PRIOR FILING DATE: 2000-04-05
; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5983

; LENGTH: 320
; TYPE: PRT

; ORGANISM: Proteus mirabilis
US-09-543-681A-5983

Alignment Scores:
Pred. No.: Length: 9,5e-48 320

Score: 491.50 Matches: 114
Percent Similarity: 55.45% Conservative: 54

Best Local Similarity: 37.62% Mismatches: 130
Query Match: 30.91% Indels: 5

DB: Gaps: 3

US-10-049-750-11 (1-921) x US-09-543-681A-5983 (1-320)

QY 7 ATCCGGTTTATGGCTTAACATGTTGGACCTTATCACCTACCAACACAGATGCCCAA 66
Db 18 LeuAlaValLeuGlySerIleAsnValAspHisIleMetAsnIleAlaGlnPheProLys 37

QY 67 GAAGGGAAACTCTGAAGCGCGGCTTTAAATCGGCTCGCGGGAAGGGCGAAC 126
Db 38 ProGlyGluThrValIleGlyHisAspTyrLysIleAlaPheGlyGlyLysGlyAlaAsn 57

QY 127 CAGGCGTGGCGCGCTCAAGCTCAATTCAAAAGTATTCATGTTGACCAAGTGGCGAC 186
Db 58 GlnAlaValAlaCysGlyArgSerGlyAlaAspIleThrPheIleAlaCysValGlyAsp 77

QY 187 GATATTTTTCGCGACCAACACCATTCGTAATCTCGAATCTGGGGATCAATACGACGTAT 246
Db 78 AspAlaIleGlyArgGluIleAlaGlnLeuLysThrAspAsnIleAspThrAspAla 97

QY 247 GTAGAAAAGTACCGGTACACGACGCGGTAGCGCGGATTTTGTCAACGCCCACTCC 306
Db 98 IleArgIleIleProLysThrProThrGlyValAlaMetIleLeuValAsnGluGlnGly 117

QY 307 ACCAACAGCATTCGTATCATCAAGCGCTAACAGTTCTCTCGCGGGAAGATATCGAT 366
Db 118 GluAsnValIleSerIleValAlaGlyAlaAsnSerAlaLeuThrProSerHisLeuHis 137

QY 367 CGCGCGGCGAAGATTAAATAAATGCCAGCTTATTTCTGCACTGGAAGTTACGCTT 426
Db 138 GlnTyrArgHisIleIleGluGlnAlaAspAlaLeuLeuMetGlnLeuGluSerProLeu 157

QY 427 GAAACGGTTTATCACCAATAGAATTGGCAAGAACACCGGATTTGAAGTGTATTAAC 486
Db 158 AspThrValPheGluAlaAlaLysGlnAlaLysAlaHisGlnThrLysValIleLeuAsn 177

QY 487 CTGCGCCAGCATTCAGGAATTAGATATGCTTATGCTGCTAAATCGCATTTCTTTGA 546
Db 178 ProAlaProAla---GlnProLeuSerAspGluPheLeuSerPheIleAspIleThr 196

QY 547 CTAATAGAAACCGAGCTGGAAATATTAACCGGTATGCCAGTGATACCTATGACCATATT 606
Db 197 ProAsnGluThrGluAlaGluIleLeuThrGlyIleSerVal-----HisAspGluVal 214
QY 607 CGCGCA-----CGCGCACGTTTCGGTGTAGATAAAGGCTGAACAATATTATTGTACC 660
Db 215 GlyAlaAlaLysAlaAlaAsnIleLeuHisSerLysGlyIleHisValLeuIleThr 234
QY 661 ATGGCGGAGAAAGCGCGCTGTGGATGACGCGTGACCGACGAGAAAGTCCATGTCGGCGTT 720
Db 235 LeuGlySerArgGlyValTrpPheSerGluGlnGlyThrGlyMetIleIleProGlyPhe 254
QY 721 AGAGTGAAGCTGTTGATACACGCGCGCGCGGATGCTTATCGGCTGTTTCGGCGCAT 780
Db 255 ArgValGluAlaValAspThrIleAlaAlaGlyAspThrPheAsnGlyAlaPheValThr 274
QY 781 TACTACGTCCAGACGCGGGATCTGGAAGCGCGCATGAAAGAAAGCCGCTCTTTGCGCGCT 840
Db 275 AlaIleLeuGluGlyLysSerAlaHisAspAlaIleArgPheAlaHisAlaAlaAla 294
QY 841 TTCAGCGTCACCGGGAAGGCAACCAATCTCTTATCCAGCATTTGAGCAATTTAATGAG 900
Db 295 IleAlaValThrArgHisGlyAlaGlnSerSerValProTrpArgAspGluIleLysSer 314
QY 901 TATCTTTTCG 909
Db 315 PheLeuAla 317

RESULT 4

US-09-710-279-3062
; Sequence 3062, Application US/09710279
; Patent No. 6703492

GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN

; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS48005

; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09

; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09

; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3062
; LENGTH: 302

; TYPE: PRT
; ORGANISM: Artificial Sequence

FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic

OTHER INFORMATION: amino acid sequence
US-09-710-279-3062

Alignment Scores:
Pred. No.: Length: 6,98e-44 302

Score: 458.00 Matches: 108
Percent Similarity: 55.96% Conservative: 61

Best Local Similarity: 35.76% Mismatches: 129
Query Match: 28.81% Indels: 4

DB: Gaps: 4

US-10-049-750-11 (1-921) x US-09-710-279-3062 (1-302)

QY 7 ATCCGGTTTATGGCTTAACATGTTGGACCTTATCACCTACCAACACAGATGCCCAA 66
Db 1 ValIleValIleGlySerThrAsnValAspLysPheLeuAsnValLysArgPheProLys 20

QY 67 GAAGGGGAAACTCTGGAAGCG---CGGCGGTTTAAATCGGCTGCGCGGGAAGGGCG 123
Db 21 ProGlyGluThrLeuHisIleAsnGlnAlaGlnLysGluPheGlyGlyLysGlyAla 40

QY 124 AACAGGCGGTGGCGCGCTCAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCG 183
Db 41 AsnGlnAlaIleAlaAlaSerArgLeuAlaAlaAspThrThrPheIleSerLysValGly 60

```
QY 184 GACGATATTTTCCGCAACACCACTTCGTAACTCTCGAATCTCGGGGATCAATACGACG 243
Db 61 LysAspGlyAsnAlaAsnPhelLeuGluAspPheLysLeuAlaGlyLeuHisThrGln 80
QY 244 TATGTAGAAAAGTAGTACCGGTACCGACGCGCGGTAGCGCCGATTTTCGTCAACGCCAAC 303
Db 81 TyrIleLeuThrSerGluSerGluGluThrGlyGlnAlaPheIleThrValAspGluAla 100
QY 304 TCCAGCAACAGCATTTCTGATCATCAAGCGGTACAAAGTTCTCTCGCGGGAAGATATC 363
Db 101 GlyGlnAsnThrIleLeuValTyrGlyGlyAlaAsnMetThrLeuSerAlaThrAspVal 120
QY 364 GATCGCGCGCGGAGAGATTAAAAATGCCCAGCTATTGTCTTCCAACTGGAAGTTCAG 423
Db 121 GluMetSerAlaAspAlaPheIleGlyAlaAspPheValValAlaGlnLeuGluValPro 140
QY 424 CTTGAAACGGTTTATCACCAATAGAAATTTGGCAAGAAACACGGGATGGAAGTTATTA 483
Db 141 PheGluAlaIleGluGlnAlaPheLysIleAlaArgLysGlnAsnIleThrThrValLeu 160
QY 484 AACCTGCGCCAGCATTCGGGAATTAGATATGCTTATGCTCTTAATGCTTAATGCGATTTCTTT 543
Db 161 AsnProAlaProAlaIle---GluLeuProLysSerLeuLeuGluLeuThrAspIleIle 179
QY 544 GTACCTAATGAACCGAGCTGGAAATATTAAACGGGTATGCCAGTGGATACCTATGACCAT 603
Db 180 IleProAsnGluThrGluAlaGluLeuLeuThrGlyIleSerIleAsnAsnGluSerAsp 199
QY 604 ATTCCGCGACGGCGCTTCGCTGTGTAGATAAGGCTGAAACAATATTATTGTCCACCATG 663
Db 200 MetLysGluThrAlaThrTyrPheLeuAspLeuGlyIleSerAlaValleuIleThrLeu 219
QY 664 GCGGAGAAAGCGCGCTGTGGATGACGCTGACAGGAAGTCCATGTTCCGCGGTTTAGA 723
Db 220 GlyGluGlnGlyThrTyrCysAlaTyrGlnGluGlnTyrLysMetIleProAlaCysAsn 239
QY 724 GTGACGCTGTGTATACACGCGCGCGGGATGCTTATCGCTGTTC---GCGCAT 780
Db 240 ValLysAlaIleAspThrThrAlaAlaGlyAspThrPheIleGlyAlaPheLeuSerGlu 259
QY 781 TACTAGTCACAGCGGGATGTGGAAGCGGCATGAAAGCGCGCTCTCTTTCGCGCT 840
Db 260 LeuAsnLysAspLeuSerAsnIleGluSerAlaIleArgLeuAlaAsnGlnAlaSerSer 279
QY 841 TTCAGCGTCACCGGGAAGGCCCAATCTCTTATCCAAGCATTCGACCAATTTAAT--- 897
Db 280 LeuThrValcIlnArgLysGlyAlaGlnAlaSerIleProThrArgLysgluValGluAla 299
QY 898 GAGTAT 903
Db 300 GluTyr 301
```

RESULT 5

```
US-09-489-039A-14043
; Sequence 14043, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Berton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 14043
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-14043
```

```
Alignment Scores:
Pred. No.: 3,19e-43 Length: 330
Score: 452.50 Matches: 111
Percent Similarity: 53.31% Conservative: 50
Best Local Similarity: 36.75% Mismatches: 138
Query Match: 28.46% Indels: 3
DB: 2 Gaps: 3
```

US-10-049-750-11 (1-921) x US-09-489-039A-14043 (1-330)

```
QY 7 ATCCGGGTATTTGGCTCTAAACATGTGGACCTTATCACCTACACCAACAGATGCCCAA 66
Db 28 LeuValValLeuGlySerIleAsnAlaAspHisIleLeuAsnLeuAspAlaPheProThr 47
QY 67 GAAGGGGAAACTCTGAAGCGCGCGTCTAAATCGGCTGGCGGAAAGGGGCGAAC 126
Db 48 ProGlyGluThrValThrGlyHisIstyrGlnValAlaPheGlyGlyLysGlyAlaAsn 67
QY 127 CAGGCGGTGGCGCGCTAAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGAC 186
Db 68 GlnAlaValAlaAlaGlyArgSerGlyAlaAspIleAlaPheIleAlaCysThrGlyAsp 87
QY 187 GATATTTTTCGCAACACCACTTCGTAACTCTCGAATCTCTGGGATCAATACGCGTAT 246
Db 88 AspAspIleGlyGluArgIleArgGlnLeuAlaSerAspLysIleAspValAlaPro 107
QY 247 GTAGAAAAGTACCTGTACACGACGCGGTAGCGCGATTTTCGTCAACGCCCACTCC 306
Db 108 ValArgAlaValAlaGlyGluAlaThrGlyValAlaLeuIlePheValAsnAlaGluGly 127
QY 307 AGCAACAGCATTTCTCATCATCAAGCGCTAAACAAGTTTCTCTCGCGGAAGATTCGAT 366
Db 128 GluAsnValIleGlyIleHisAlaGlyAlaAsnAlaAlaLeuSerValSerGlnValGlu 147
QY 367 CGCGCGGGAAGATTTAAAAATGCGCATTTATGTTCTGCAACTGGAAGTTCAGTT 426
Db 148 AlaGluLysGluArgIleAlaSerAlaGlnAlaLeuLeuMetGlnLeuSerProLeu 167
QY 427 GAAACGGTTTATCAGCAATAGAAATTTGCGAAGAACACGCGGATTTGAAGTGTATTAAAC 486
Db 168 GluSerValIleAlaAlaLysIleAlaHisHisHisThrThrValValLeuAsn 187
QY 487 CTCTGCGCAGCATTCGGGAATTAGATATGCTTATGCTGTAAATGCGATTTCTTTGTA 546
Db 188 ProAlaProAla---ArgGluLeuProAspGluLeuLeuAlaLeuValAspIleThr 206
QY 547 CTAATGAAACCGAGCTGGAATATTACCGGTATGCCAGTGCATACCTATGACCATATT 606
Db 207 ProAsnGluThrGluAlaGluLysLeuThrGlyIleArgValGluSerAspGluAspAla 226
QY 607 CGCGCAGCGGCACGTTTCGCTGTAGATAAGCGCTGAACAATATTATTCTCACCATTGGC 666
Db 227 AlaLysAlaAlaAspValLeuHisAlaLysGlyIleGlyThrValMetIleThrLeuGly 246
QY 667 GAGAAAGCGCGCTGTGGATGACGGGTGACCAAGGAAGTC---CATGTTCCGGCGTTTGA 723
Db 247 SerArgGly---ValTrpLeuSerAlaGluGlyGluSerArgArgIleProGlyPheArg 265
QY 724 GTGAACGCTGTGTATACCAAGCGCGCGCGATGCTTATCGCTGTGTTTCGCGCATTAC 783
Db 266 ValGlnAlaIleAspThrIleAlaAlaGlyAspThrPheAsnGlyAlaLeuValThrAla 285
QY 784 TACTGCCAGCGGGGATGTGGAAGCGGCATGAAAGCGCTCTCTTTCGCGCTTTC 843
Db 286 LeuLeuGluGlyThrAlaLeuProGluAlaIleArgPheAlaHisAlaAlaIle 305
QY 844 AGCGTCACCGGGAAGGCCAACCTCTTATCCAAGCATTCGACCAATTAATAGTAT 903
Db 306 AlaValThrArgLysGlyAlaGlnProSerValProTrpArgThrGluIleAspGluPhe 325
QY 904 CTTTCG 909
Db 326 LeuAla 327
```

```

RESULT 6
US-09-134-001C-4521
; Sequence 4521, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4521
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4521

Alignment Scores:
Pred. No.: 3,55e-43 Length: 315
Score: 452.00 Matches: 107
Percent Similarity: 55.63% Conservative: 61
Best Local Similarity: 35.43% Mismatches: 130
Query Match: 28.43% Indels: 4
DB: 2 Gaps: 4

US-10-049-750-11 (1-921) x US-09-134-001C-4521 (1-315)
QY 7 ATCCGGTATTGGCTTAACATGGTGGACCTTATCACCTACCAACACAGATGCCCAA 66
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
14 ValIleValIleGlySerThrAsnValAspLysPheLeuAsnValIysArgPheProLys 33
QY 67 GAAGGGGAAACTCTGGAAGCG---CCGGCGTTTAAATCGGCTGCGCGGAAAGGGCG 123
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 ProGlyLeuThrLeuHisIleAsnGlnAlaGlnLysGluPheGlyGlyGlyAla 53
QY 124 ACCAGCGCGTGGCGCCCTAAGCTCAATTCAAAGTATTGATGTTGACCAAGTGGC 183
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
54 AsnGlnAlaIleAlaAlaSerArgLeuAlaAspThrPheIleSerLysValGly 73
QY 184 GAGCATATTTTGGCCGACCAACCATTCGTAATCTCGAATCTCGGGGATCAATACGACG 243
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
74 LysAspGlyAsnAlaAsnPhelleuGluAspPheLysLysAlaGlyIleHisThrGln 93
QY 244 TATGTAGAAAAAGTACCGTGTACACAGCGCGGTAGCGCGATTTTCTCAACGCCAAC 303
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
94 TyrIleLeuThrSerGluSerGluThrGlyGlnAlaPheIleThrValAspGluAla 113
QY 304 TCAGCAACAGATTTCTGATCATCAAGCGGTAAAGTTTCTCTCGCGGGAAGTATC 363
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
114 GlyGlnAsnThrIleLeuValTyrGlyAlaAsnMetThrLeuSerAlaThrAspVal 133
QY 364 GATCGCGCGCGGAACATTTAAAAAATCGCAGTTATTGTTCTGCAACTGGAAGTTGAG 423
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 GluMetSerValAspAlaPheIleGlyAlaAspPheValAlaGlnLeuGluValPro 153
QY 424 CTTGAACCGTTTATCACGCAATAGAAATTTGGCAAGAACACACGGGATGAAGTTGTTATTA 483
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154 PheGluAlaIleGluGlnAlaPheLysIleAlaAatGlyGlnAsnIleThrThrValLeu 173
QY 484 ACCCTGCGCGCAGCATTCAGGAATAGATATGCTTATGCTGCTGTAATTCGATTTCTTT 543
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
174 AsnProAlaProAlaIle---GluLeuProLysSerLeuLeuLeuThrAspIleIle 192
QY 544 GTACCTAATGAACCGAGCTGGAATATTAAACCGGTATCCAGTGGATACCTATGACCAT 603
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
193 IleProAsnGluThrGluAlaGluLeuLeuThrGlyIleSerIleAsnAsnGluSerAsp 212
```

```

QY 604 ATTCGGCAGCGCAGCGTTCGCTGTAGATAAAGGCTCAACATATATTATTCACCATG 663
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
213 MetLysGluThrAlaThrPheLeuAspLeuGlyIleSerAlaValLeuIleThrLeu 232
QY 664 GCGAGAAAGCGCGCTGTGGATGACGCTGACCGAGAGTCCATGTCGGCGGTTTGA 723
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
233 GlyGluGlnGlyThrTyrCysAlaTyrGlnGluGlnTyrLysMetIleProAlaCysAsn 252
QY 724 GTCAACGCTGTTGATACACAGCGCGCGCGATGCTTTATCGGCTGTTTC---GCGCAT 780
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
253 ValLysAlaIleAspThrThrAlaAlaGlyAspThrPheIleGlyAlaPheLeuSerGlu 272
QY 781 TACTAGCTCAGAGCGGGGATGTGNAAGCGGCATGAAAGCCCTCTCTTTCGGCT 840
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
273 LeuAsnLysAspLeuSerAsnLeuGluSerAlaIleArgLeuAlaAsnGlnAlaSerSer 292
QY 841 TTCAGCGTCACCGGAAAGGACCAATCTCTTATCCAGCATTCAGCAATTTAAT--- 897
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
293 LeuThrValGlnArgLysGlyAlaGlnAlaSerIleProThrArgLysGluValGluAla 312
QY 898 GAGTAT 903
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
313 GluTyr 314

RESULT 7
US-09-543-681A-7498
; Sequence 7498, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7498
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7498

Alignment Scores:
Pred. No.: 1.17e-42 Length: 311
Score: 447.50 Matches: 108
Percent Similarity: 55.08% Conservative: 60
Best Local Similarity: 35.41% Mismatches: 134
Query Match: 28.14% Indels: 3
DB: 2 Gaps: 3

US-10-049-750-11 (1-921) x US-09-543-681A-7498 (1-311)
QY 7 ATCCGGTATTGGCTTAACATGGTGGACCTTATCACCTACCAACACAGATGCCCAA 66
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
9 LeuValIleLeuGlySerValAsnValAspHisIleLeuAsnValAlaHisPheProHis 28
QY 67 GAAGGGGAAACTCTGGAAGCGCGGCTTTAAATCGGCTGCGCGGAAAGGGCGAAGC 126
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29 ProGlyGluThrIleSerGlyAsnGlnTyrGlnMetValPheGlyGlyLysGlyAlaAsn 48
QY 127 CAGGCGCGTGGCGCGCTAAGCTCAATTCAAAGTATTGATGTTGACCAAGTGGCGAC 186
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
49 GlnAlaValAlaAlaGlyArgCysGlyAlaAsnIleThrPheLeuAlaCysLeuGlyAsn 68
QY 187 GATATTTTTCGCGACACACCATTCGTAATCTCGAATCTCGGGGATCAATACGAGTAT 246
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
69 AspAspIleGlyGlnSerAlaLysAlaGlnLeuIleThrAspLysIleAspThrAspCys 88
QY 247 GTAGAAAAGTAGTACCGTGTACAGCGCGGTAGCGCGATTTTCGTCAACGCCACTCC 306
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
89 IleGluLeuIleAspGluAlaThrGlyValAlaIlePheValAsnGlnGlnGly 108
```

307 AGCAACAGCATTCTGATCATCAAGCGCTAAACAAGTTTCTCTCGCGGAGATATCGAT 366
 109 GluAenValIleGlyIleHisAlaGlyAlaAenGlyArgLeuAenArgGlyTyrValGlu 128
 367 CGCGCGGCGGAGATTAAATAAATGCCAGCTATTATTGTTCTGCAACTGGAAGTTCCAGCTT 426
 129 ArgTyrAenSerValIleLysLysAlaAenAlaLeuLeuMetGlnLeuGluSerProLeu 148
 427 GAAACGGTTTATCAGCAATAGAAATTTGGCAGAAACACCGGATTTGAAGTTGTTATAAC 486
 149 AspSerValLeuLysAlaAlaGlnValAlaLysGlnGluSerValGlnValIleLeuAen 168
 487 CTTGCGCCAGCATTAAGGAAATTTAGATATGTCTTATGCTGTAATGCGATTCTTTGTA 546
 169 ProAlaProAla---GlnAlaLeuProAepGluLeuLeuSerLeuValAspIleLeuThr 187
 547 CTTAATGAAACCGAGCTGGAATATTTAACCGGTATGCCAGTGGATACCTATGACCATATT 606
 188 ProAenGluThrGluThrGluTyrLeuThrGlyValLysValIleAepAspGluSerAla 207
 607 CGCGCAGCGGCGAGTTGCTGTGTAGATAAAGCGGTGAACAATATTATTGTCACCATGGGC 666
 208 GlnArgAlaSerAenIleLeuHisAspLysGlyIleGluThrValLeuIleThrLeuGly 227
 667 GAGAAAGCGCGCTGGTATGACGGT---GACCAGGAAGTCCATGTTCCGGCGTTTGA 723
 228 SerArgGly---ValTrpValSerGlnAenAenGlnGlyThrMetValProAlaPheLys 246
 724 GTGAACGCTGTTGATACACAGCGCGCGCGATGCTTTATCGGCTGTTTCGCGCATTTAC 783
 247 ValThrAlaIleAspThrIleAlaAlaGlyAspThrPheAenGlyAlaPheIleThrAla 266
 784 TACGTCCAGAGCGGGAGTGTGAAGCGCCCATGAAAGCGCTCTCTTTCGCGCTTTC 843
 267 LeuLeuGluGlyLeuProMetMetLysAlaIleLysPheAlaHisAlaAlaAlaIle 286
 844 ACCTCACCAGGAAAGGACCAATCTCTTATCCAAAGCATTTGACCAATTTAATGAGTAT 903
 287 AlaValThrArgAlaGlyAlaGlnProSerValProTrpArgHisGluIleGluAlaPhe 306
 904 CTTTCGTTGAACGAA 918
 307 LeuSerGlnHisAsp 311

RESULT 8
 US-09-543-681A-8124
 ; Sequence 8124, Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:
 ; APPLICANT: GARY BRETON
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 ; FILE REFERENCE: 2709.1002-001
 ; CURRENT APPLICATION NUMBER: US/09/543,681A
 ; PRIORITY FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: US 60/128,706
 ; NUMBER OF SEQ ID NOS: 8344
 ; SEQ ID NO 8124
 ; LENGTH: 406
 ; TYPE: PRT
 ; ORGANISM: Proteus mirabilis
 US-09-543-681A-8124

Alignment Scores:
 Pred. No.: 1,85e-39 Length: 406
 Score: 420.50 Matches: 98
 Percent Similarity: 54.43% Conservative: 68
 Best Local Similarity: 32.13% Mismatches: 134
 Query Match: 26.45% Indels: 5
 DB: 2 Gaps: 3

US-10-049-750-11 (1-921) x US-09-543-681A-8124 (1-406)
 QY 7 ATCCGGGTATTGGCTCTAAACATGCTGACCTTATCACCTACCAACACAGATGCCCAA 66
 DB ValCysValLeuGlySerPheAenValAspMetIleSerTyrLeuProArgLeuProGlu 113
 QY 67 GAAGGGGAAACTCTGGAAGCGCGCGTTTAAATCGGCTCGCGGCGGAAAAGGGCGAAG 126
 DB ThrGluGluSerLeuAlaAenLysPheIlePheSerProGlyGlyGlyCysAen 133
 QY 127 CAGCGCTGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGAC 186
 DB GlnAlaLeuAlaSerTyrAlaAenValHisPheIleThrLysIleGlyThr 153
 QY 187 GATATTTTTCGCGACACACCATTCGTAATCTCGAATCTGGGATCAATACGACGATAT 246
 DB AspGlnPheSerAspTyrAlaValAenPheIleSerSerArgIleLysThrSerThr 173
 QY 247 GTAGAAAAGTACCGTGTACACAGCAGCGCGGTAGCGCGGATTTTCGTCAACGCCAATCC 306
 DB IleTyrGlnThrGluAenHisGlnThrGlyThrAlaSerIlePheValSerGluGluSer 193
 QY 307 AGCAACAGCATTCG---ATCATCAAGCGCTAAACAAGTTTCTCTCGCGGGAAGATATC 363
 DB GlyGluAenIleIleSerIleTyrSerGlyAlaAenMetAspIleSerSerAspGluVal 213
 QY 364 GATCCGCGCGGGAAGATTTAAAAAATGCCAGCTATTGTTCTGCAACTGGAAGTTGAG 423
 DB LysIleGlnLysAspLysIleIleAspAlaAspIleIleLeuLeuGlnLeuGluThrAen 233
 QY 424 CTTGAAACGGTTTATCACGCCAATAGAAATTTGGCAAGAAACACGGGATTTGAAGTTATTA 483
 DB IleGluAlaLeuLysGluIleIleSerIleGlyAenGluAenAsnIleProIleLeuLeu 253
 QY 484 AACCTCGCGCCAGCATTTACGGGAATTAATGATATGCTTATGCTGTAATGCGATTCTTT 543
 DB AenProAlaProTyrAsnLysIleIleAsp---LysLeuLeuProMetLeuAspIleLeu 272
 QY 544 GTACCTATGAACCGCGCTGGAATATTTAACCGGTATGCCAGTGTGATACCTATGACCAT 603
 DB ThrProAenGluThrGluAlaSerLeuLeuSerGlyIleGluValValAspLeuValSer 292
 QY 604 ATTCGCGCAGCGCACGCTTCGCTGTAGATAAAGCGCTGAAACAATATTATTTCACCATG 663
 DB AlaLysAenAlaAlaAenAlaIleTyrGlnLysGlyValAenLysValIleThrLeu 312
 QY 664 GCGGAGAAAGCGCGCTGTGTGATGACGCTGACAGGAAGTCCATGTTCCGCGGTTAGA 723
 DB GlySerLysGlySerLeuAlaTyrAspGlyTyrLysTyrIleTyrSerProAlaTyrPro 332
 QY 724 GTGAACGCTGTTGATACACAGCGCGCGGATGCTTTATCGGCTGTTTCGCGCATTTAC 783
 DB AlaValValLysAenThrAlaGlyAlaGlyAepAlaPheAenGlyAlaLeuAlaAenSer 352
 QY 784 TACGTCCAGAGCGGGATGTGAAGCGCGCATGAAAGCGCTCTCTTTCGCGCTTTC 843
 DB LeuAlaLysGlyLysGlnPheSerTyrAlaLeuArgTyrAlaSerAlaPheAlaSerLeu 372
 QY 844 ACCGTCCAGGAAAGGACCAATCTCTTATCCAAAGCATTTGACCAATTTAATGAGTAT 903
 DB AlaValGlu-----ThrSerAenAlaSerGluMetProGluAspIleAenValMet 389
 QY 904 CTTTCGTTGAACGAA 918
 DB HisArgIleAenGln 394

RESULT 9
 US-09-489-039A-9668
 ; Sequence 9668, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 9668
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-9668

Alignment Scores:
Pred. No.: 1,116-35 Length: 342
Score: 387.50 Matches: 103
Percent Similarity: 50.16% Conservatives: 50
Best Local Similarity: 33.77% Mismatches: 143
Query Match: 24.37% Indels: 9
DB: 2 Gaps: 4

US-10-049-750-11 (1-921) x US-09-489-039A-9668 (1-342)

QY 7 ATCCGGTTATTGGCTCTAACATGGTGGACCTTATCACCTACACCAACAGATGCCCAA 66
Db 34 ValCysValPheGlySerPheAsnPheAspMetValAlaArgValAspArgPheProVal 53
QY 67 GAAGCGAAACTCTGAAGCGCGCGCTTAAATCGGCTCGCGGAAAGGGCGGAAC 126
Db 54 ProGlyGluSerLeuValAlaCysGlySerMetThrSerAlaGlyGlyysGlyAlaAsn 73
QY 127 CAGCGCGTGGCGCCCTAGCTCAATTCAAAGTATTGATGTGACCAAGTGGCGAC 186
Db 74 GlnAlaThrAlaAlaLeuLysAlaGlyAlaAsnValHisTyrIleGlyLysIleGlyAsn 93
QY 187 GATATTTTCCGACACACCATTCGTAATCTCGAATCTGGGATCAATACGAGTAT 246
Db 94 AspThrPheGlyHisPheAlaArgArgHisLeuLysGlyValGlyPheAsnAlaValThr 113
QY 247 GTAGAAAAAGTACCGTGTACCGACGCGCGTGTAGCGCGATTTTCGTCAACGCCAAC--- 303
Db 114 LeuLeuValAlaGluThrProThrGlyAsnAlaLeuIleTyrValAlaGlyAsnAsp 133
QY 304 TCCAGCAACAGCATTCGTATCATCAAGGCGCTAACAAAGTTCTCTCGCGGGAAGATATC 363
Db 134 AlaGluAsnMetIleAlaValAspProGlyAlaAsnMetThrValThrAspAspGluIle 153
QY 364 GATCGCGCGGGAAGATTTAAAAATGCGCCTTATTGTTCTGCAACTGGAAGTTCAG 423
Db 154 AlaGlyCysIleProAlaIleGlyCysAlaAspValValLeuValGlnLeuGluAsnAsn 173
QY 424 CTTGAAACCGTTTATCACGCAATAGAAATTTGGCAAGAACACACGGGATTGAAGTGTATTAT 483
Db 174 LeuSerAlaIleGluGlnValIleAspAlaGlyLysGlnAlaGlyAlaLeuValIleLeu 193
QY 484 AACCTCGCGCAGCATTCACGGGAATTAGATATGCTTATGCTTCCTGTAATGCGGATTTCTTT 543
Db 194 AsnProAlaPro---TrpGlnProValGluHisAlaLeuLeuArgLysValAspLeuLeu 212
QY 544 GTACCTAATCAACCGCATTCGGAATATTAACCGGTATGCCAGTGGATACCTATGACCAT 603
Db 213 ThrProAsnAlaThrGluAlaGlyLeuMetThrGlyArgValAspSerLeuThrAla 232
QY 604 ATTCCGCGACCGCGCTTCGCTGTAGATAAGGCTGAACAATATTATTGTCACCATG 663
Db 233 AlaAlaGluAlaAlaAspValLeuHisAlaGlnGlyAlaArgAsnValIleIleThrLeu 252
QY 664 GCGGAGAAAGCGCGCTGTGGATGACGCGTACAGGAAGTCCATGTTCCGCGGTTTAGA 723
Db 253 GlyAlaSerGlyAlaLeuLeuSerGluHisGlyValLysSerProIleProCysPhePro 272
QY 724 GTGAACGCTGTGATACCAAGCGCGCGCGATGCTTTATCGCTGTGTTCCGGCATTTAC 783
::: |||||:::|||||

Db 273 SerHisProArgAspThrThrGlyAlaGlyAspAlaPheAsnGlyAlaLeuAlaAArg 292
QY 784 TAGCTCCAGAGCGGGATGTGAAGCGCCATGAAAGAGCGCTCTCTTTGCCGCTTTC 843
Db 293 LeuAlaCysGlyGluProLeuGlnAlaAlaAArgPheAlaAlaAlaVal 312
QY 844 AGCTCACCAGGGAAGGACCCCAATCCTCTTATCCAAGCATTGACGCAATTAATGAGTAT 903
Db 313 SerValGluLysGlnGlyAla---SerSerLeuPro-----GluTyr 325
QY 904 CTTTCGTTGAAGCAA 918
Db 326 LeuGluAlaGlnGlu 330

RESULT 10

US-09-489-039A-7290

; Sequence 7290, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 7290

; LENGTH: 328

; TYPE: PRT

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-7290

Alignment Scores:

Pred. No.: 1,566-34 Length: 328

Score: 377.50 Matches: 99

Percent Similarity: 51.16% Conservatives: 56

Best Local Similarity: 32.67% Mismatches: 141

Query Match: 23.74% Indels: 7

DB: 2 Gaps: 5

US-10-049-750-11 (1-921) x US-09-489-039A-7290 (1-328)

QY 4 GATATCGCGTTATTGGCTCTAACATGGTGGACCTTATCACCTACACCAACAGATGCC 63
Db 24 AspValValIleGlySerLeuAsnTyrAspIleLeuValGlnGlnAspArgLeuPro 43
QY 64 AAAGAAGGGGAACTCTGAAGCGCGCGCTTAAATCGGCTCGCGGGAAGGGCGG 123
Db 44 GluLeuGlyGluThrPheThrGlyAsnGluLeuMetLeuMetProGlyGlyysGlyAla 63
QY 124 AACCAAGCGCGTGGCGCGCTTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCG 183
Db 64 AsnGlnAlaValGlnCysSerArgLeuGlyLeuAsnValSerMetValGlyCysValGly 83
QY 184 GACGATATTTTCCGACCAACACCATTCGTATCTCGAATCTCTGGGGGATCAATACGACG 243
Db 84 AsnAspIleTyrGlySerGluLeuIleAsnSerLeuArgGluAsnValSerValGlu 103
QY 244 TATGTAGAAAAGTACCGGTATACCGACGCGCGGTAGCGCGCATTTTCGTCAACGCCAAC 303
Db 104 AsnValAsnLys---ArgGlyThrThrGlyIleGlyIleValGlnIleLeuGluSer 122
QY 304 TCCAGCAACAGCATTCGTATCATCAAGGCGCTGAACAAGTTCTCTCGCGGGAAGATATC 363
Db 123 GlyAspTyrCysSerThrIleIleLysGlyAlaAsnTyrLeuIleSerGluAspAspIle 142
QY 364 GATCGCGCGCGGAAGATTTAAAAAATGCCAG---CTTATTGTTCTGCAACTGGAAGTT 420
Db 143 -----ThrGluSerLeuPheGluGlyGlnProLeuValIleLeuGlnSerGluIle 159
QY 421 CAGCTTGAACCGCTTTATCAGCAATAGAAATTTGGCAAGAACACACGGGATGAGTGTTA 480

Db 160 ProAlaProValValGluTyrIleIleGlyValAlaSerThrHisHisCysArgIleIle 179
| | | | | : : : | | | : : : : :
Qy 481 TTAACCCCTGCCAGCATTACGGAAATTAGATATGCTTATGCTGTAAATGCCATTTC 540
| | | | | : : : : :
Db 180 LeuAsnAsnAlaProAla---ArgAspValSerAlaHisAlaLeuSerLeuValAspTyr 198
| | | | | : : : : :
Qy 541 TTGTACCTAANTGAACCGAGCTGGNAATATTAAACCGGTATGCCAGTGGATACCTATGAC 600
| | | | | : : : : :
Db 199 LeuValValAsnGluThrGluAlaAlaPheMetSerGlyAlaAspValSerSerIleAsp 218
| | | | | : : : : :
Qy 601 CATATTGGCGCGCGCGCTTCGCTGGTGTAGATAAAGCGCTGAACAAATATTATGTCACC 660
| | | | | : : : : :
Db 219 AspAlaHisSerCysAlaThrGlyLeuHisLysArgValLysGlyGlnValIleIleThr 238
| | | | | : : : : :
Qy 661 ATGGCGGAGAAAGCGCGCTGTGGATAGCAGCGTACAGGAAGTCCATGTCCTCCGCG--- 717
| | | | | : : : : :
Db 239 LeuGlyGluLysGlyAlaValLeuSerGlnGluHisGlyThrArgHisPheProAlaVal 258
| | | | | : : : : :
Qy 718 TTTAGAGTGAACGCTGTTATATACAGCGCGCGGGGATGCTTATCGGCTGTTTCGG 777
| | | | | : : : : :
Db 259 PheCysProAspValValAspThrThrGlyAlaGlyAspSerPheIleGlyGlyLeAla 278
| | | | | : : : : :
Qy 778 CATTAAGTACGTCAGAGCGCGGATGTGAAGCGGCATCAAAAAAGCCGCTCTCTTTGCC 837
| | | | | : : : : :
Db 279 TyrCysIleValAsnGlyPheSerLeuThrAspAlaIleProPheAlaAlaGluIleSer 298
| | | | | : : : : :
Qy 838 GTTTTCAGCGTCACCGGGAAGCCCAATCTCTTATCCAGCATTCAGCAATTTAAT 897
| | | | | : : : : :
Db 299 SerCysSerIleGlnLysTyrGlyGlyGlnAsnSerPheProMetLeuProAspValSer 318
| | | | | : : : : :
Qy 898 GAGTATCTT 906
| | |
Db 319 HisAlaLeu 321

RESULT 11

US-09-134-000C-4524
; Sequence 4524, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4524
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-4524

Alignment Scores:

Pred. No.: 2,72e-34 Length: 529
Score: 376.50 Matches: 97
Percent Similarity: 49.51% Conservative: 54
Best Local Similarity: 31.80% Mismatches: 143
Query Match: 23.68% Indels: 11
DB: 2 Gaps: 3

US-10-049-750-11 (1-921) x US-09-134-000C-4524 (1-529)

Qy 13 GTTATTGGCTCTAACATGTGGACCTTATCACCTACCAACCCAGATGCCAAGAAGGG 72
| | | | | : : : : :
Db 221 ValValGlySerLeuAsnValAspAspTyrLeuTyrSerThrAsnLeuProHisAsnGly 240
| | | | | : : : : :
Qy 73 GAAACTCTGGAAGCGCGCGGTTTAAATCGGCTGCGCGGGAAGAGCGGACACGAGGCC 132
| | | | | : : : : :
Db 241 LysThrAsnPheLeuSerSerTyrAlaLysPheProGlyGlyLysGlyLeuAsnGlnAla 260
| | | | | : : : : :

Qy 133 GTGCGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAAAGTGGCGACGATATT 192
| | | | | : : : : :
Db 261 ValGlyLeuThrLysLeuGlyHisGlnAlaThrLeuIleGlyCysLeuGlySerAspThr 280
| | | | | : : : : :
Qy 193 TTTGCCGCAACACCATTCGTAATCTCGAATCCTCGGGGATCAATACGAGTATGTAGAA 252
| | | | | : : : : :
Db 281 AspAlaAsnTyrLeuTyrLysGluLeuGlyLysTyrHisValThrThrAspGlyIleThr 300
| | | | | : : : : :
Qy 253 AAAGTACCGTGTACAGACGCGGTAGCGCCGATTTCTGTCGCGGAAGATTCGTAACGCCAATCCAGCAAC 312
| | | | | : : : : :
Db 301 ArgIleGlnAspThrGluThrGlyGlnAlaTyrIleTyrValGluThrSerGlyAspSer 320
| | | | | : : : : :
Qy 313 AGCATTCGATCATCAAAAGCGCTAACAAAGTTTCTCTCGCGGAAGATTCGATCGCGCG 372
| | | | | : : : : :
Db 321 MetIleSerIleLeuProGlyAlaAsnThrAlaLeuThrProLysLysIleAlaGlnGln 340
| | | | | : : : : :
Qy 373 GCGGAAGATTAAAAAATGCCAGCTTATTGTTCTGCAACTCGAAGTTCGAGTTGAAACG 432
| | | | | : : : : :
Db 341 LysHisLeuPheMetAspAlaSerPheCysLeuIleGlnThrGluIleProLeuSerAla 360
| | | | | : : : : :
Qy 433 GTTTATCAGCAATAGAAATTTGGCAAGAAACACGGGATGGAAGTGTATTAAACCTGCG 492
| | | | | : : : : :
Db 361 ValGluLysAlaCysGluIleAlaGlnHisSerGlyValProIleIleLeuLysProAla 380
| | | | | : : : : :
Qy 493 CCAGCATTCACGGGAATTAGATATGCTTATGCTGCTAAATGCGATTCTTTGTACCTAAT 552
| | | | | : : : : :
Db 381 ---AlaIleHisIleProValAsnIleLeuGluLysValAspPheIleProAsn 399
| | | | | : : : : :
Qy 553 GAAACCGAGCTGGAATATTA-----ACCGGTATGCCAGTGGATACCTATGACCAT 603
| | | | | : : : : :
Db 400 GluAspGluLeuLeuGluLeuGlnProAspThrGlyThr----- 412
| | | | | : : : : :
Qy 604 ATTCGCGCAGCGCGCTGCTGCTGTAGATAAAGGCTCAACAATATTATTCACCATG 663
| | | | | : : : : :
Db 413 LeuGluGluLysAlaAlaTyrPheLeuGluMetGlyValLysAsnValIleValThrLeu 432
| | | | | : : : : :
Qy 664 GCGGAGAAAGCGCGCTGTGGATGACGCTGACAGGAAGTCCATGCTCCGCGTTTACA 723
| | | | | : : : : :
Db 433 GlyLysGlyValLeuLeuLysThrProGlnValCysHisTyrPheProAlaThrGlu 452
| | | | | : : : : :
Qy 724 GTGAACGCTGTTGATACCAAGCGCGCGCGATGCTTTATCGGTGTTTCGCGCATTC 783
| | | | | : : : : :
Db 453 AsnIleAlaValAspSerThrGlyAlaSerAspSerPheIleSerAlaLeuAlaSerTyr 472
| | | | | : : : : :
Qy 784 TAGTCCAGAGCGGGATGTGAAGCGGCATGAAAGCCGCTCTCTTTGCGGCTTTC 843
| | | | | : : : : :
Db 473 LeuSerLysGlyTyrProThrGluAlaAlaIleGlnIleAlaIleGlnAlaGlyPhe 492
| | | | | : : : : :
Qy 844 AGCGTCACCGGGAAGGCCACCAATCTCTTATCCAGCATTCGAGCAATTTAATAGTAT 903
| | | | | : : : : :
Db 493 SerValSerLysGluGlyValIleAspSerLeuValAspHisValThrLeuGluAsnTyr 512
| | | | | : : : : :
Qy 904 CTTTCGTTGAACGAA 918
| | |
Db 513 LeuIleLysLysGlu 517
| | |

RESULT 12

US-09-543-681A-4841
; Sequence 4841, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4841
; LENGTH: 451

```
; TYPE: PRT
; ORGANISM: Proteus mirabilis
; US-09-543-681A-4841

Alignment Scores:
Pred. No.: 11e-26 Length: 451
Score: 310.50 Matches: 87
Percent Similarity: 47.90% Conservative: 61
Best Local Similarity: 28.16% Mismatches: 146
Query Match: 19.53% Indels: 15
DB: 2 Gaps: 7

US-10-049-750-11 (1-921) x US-09-543-681A-4841 (1-451)
QY 7 ATCCGGTATTGGCTTAACATGGTGGACCTTATCACCTTACCAACACGATGCCCAAA 66
Db 134 ValCysValLeuGlySerPheValValAspIleAlaAsnValAlaPheProArg 153
QY 67 GAAGGGGAAACTCTGGAAGCGCGCGGTTTAAATCGGCTGCGGGAAGGGGGAAC 126
Db 154 ValGlyGluLeuValAsnSerLysGlyAsnSerIleGlyProGlyGlyLysGlyThrAsn 173
QY 127 CAGCGCGTGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGGAC 186
Db 174 GlnAlaIleSerAlaSerPheSerAspAlaLysValHisLeuIleThrLysValGlyGlu 193
QY 187 GATATTTTGGCCGACACACCATTCGTAATCTCGAATCTCGGGGATCAAT-----ACG 240
Db 194 AspHisPheSerLysTyrAlaTyrLysTyrLeuGlnGluSerGlyIleAspSerPheThr 213
QY 241 ACGTATGTAGAAAGTACGCTGACAGCGCGGTAGCGCGGATTTTTCGTCAAGCC 300
Db 214 IlePheGlnThrGluIleGluProThrGlySerSerIleSer---TyrLeuAlaAspLys 232
QY 301 AACTCCAGCAACGATCTGATCATCAAGGCGCTAACAAGTTTCTCTCGCGGAGAT 360
Db 233 ThrGlnAsnAsnIleThrAlaThrTyrLeuGlyAlaAsnAsnThrPheThrLeuGlnGlu 252
QY 361 ATCGATCGCGCGCGGAAGATTTAAATAATGCCAGTTATTGTTCTGCAACTGGAAGTT 420
Db 253 ValAspIleSerLeuProTyrIleSerGluAlaAspValLeuLeuLeuGlnGlyGluIle 272
QY 421 CAGCTGAAACGGTTTATCAGCAATAGAAATTTGGCAAGAACACCGGATTTGAAGTGT 480
Db 273 AsnIleAspAlaAsnValLysAlaAlaSerPheAlaHisSerIleAsnLysThrValIle 292
QY 481 TTAACCTCGCGCAGCATTTACGGGATTAGATATGCTTATGCTGTAATCGGATTC 540
Db 293 LeuAsnValAlaProTyrSerAspAspLeuLysGlnLeuTyr---LeuTyrValAspPhe 311
QY 541 TTGTACCTAATGAACACGAGCTGGAATATTAAACCGGTATGCCAGTGATACCTATGAC 600
Db 312 IleThrLeuAsnAlaTyrGlnAlaSerAspTrpSerGlyIleGluIleAsnThrIleAsn 331
QY 601 CATATTCGCGCAGCGCAGCTGCTGCTGAGTAAAGGCTGAAACAATATTATGTCACC 660
Db 332 AspAlaLysGlnAlaValGluIleIleAlaGlyAsnGluLysLysLysValIleIleTyr 351
QY 661 ATGGCGGAGAAAGCGCGCTGTGGATGACGCGTGACAGGAAGTCCATGTTCCGGCGTTT 720
Db 352 IleAspGluLeuGlyValValTyrPheAspGlyArgAsnThrPheHisIleProLeu 371
QY 721 AGAGTGAACGCTTGTATACCGCGCGGGGATGCTTTATCGGCTGTTTTCGCGCAT 780
Db 372 ProSerLeuArgValAspThrMetAlaIleThrAspAlaPheAsnGlyAlaPheAlaSer 391
QY 781 TACTACGTCCAGCGGGATGTGGAAGCGCGCATGAAACCGCTCTCTCTTTCG--- 837
Db 392 LysIleAlaAlaGlyGly-----ThrMetProGluSerValLeuPheAlaSer 407
QY 838 -----GCTTTACGCTCAGCGGGAAGGACCCCATCTCTTATCCAGCAT 885
Db 408 AlaPheLeuSerAlaPhe---IleGluGlnLysGlyVal---ThrSerMetProSerLeu 425
```

```
QY 886 GAGCAATTTAATGAGTATCTTTCGTTG 912
Db 426 SerGlnValGlnAlaArgLeuLysLeu 434

RESULT 13
US-09-107-532A-3946
; Sequence 3946, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: US/09/107,532A
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneka
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3946:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 170 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...170
; SEQUENCE DESCRIPTION: SEQ ID NO: 3946:
US-09-107-532A-3946

Alignment Scores:
Pred. No.: 6.16e-20 Length: 170
Score: 250.00 Matches: 60
Percent Similarity: 53.80% Conservative: 32
Best Local Similarity: 35.09% Mismatches: 63
Query Match: 15.72% Indels: 16
DB: 2 Gaps: 4

US-10-049-750-11 (1-921) x US-09-107-532A-3946 (1-170)
QY 409 CAACGTGGAAGTTCAGCTTGAACGGTTTATCAGCAATAGATTTCGCAAGAACACGGG 468
Db 3 GlnPheGluSerArgSerAspAlaThrIleLeuAlaPheSerIleAlaLysAspAlaGly 22
QY 469 ATTGAAGTGTATTAAACCGCTGCGGACGATACGCGGAATTAGATATGCTTATGCTGT 528
Db 23 LysThrThrIleLeuAsnProAlaProAlaArgGluThrIleProThrGluLeuLeuGlu 42
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame plus n2p model

Run on: December 23, 2005, 22:42:49 ; Search time 162.356 Seconds
(without alignments)
4740.464 Million cell updates/sec

Title: US-10-049-750-11

Perfect score:

Sequence: 1 atggatatcgcggttattgg.....atcttcggtgaacgaataa 921

Scoring table: BLOSUM62

Xgapop	10.0	, Xgapext	0.5
Ygapop	10.0	, Ygapext	0.5
Fgapop	6.0	, Fgapext	7.0
Delpop	6.0	, Delext	7.0

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 3735138

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

```
-MODEL=frame_n2p_model -DEV=xlp
-Q=Cgn2/USFT0.spool_P/US10049750/runat_23122005_113555_6388/app_query.fasta_1.1742
-DB=Published Applications_AA_Main -QMT=fastcan -SUFFIX=n2p.raphm
-MINMATCH=0.1 -LOCPCL=0 -LOOPEXT=0 -UNIT=bits -START=1 -END=-1
-MATRIX=blosum62 -TRANS=human40.cd1 -LIST=45 -DocalIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MTN=0 -ALIGN=15 -MODEL=LOCAL -OUTFMT=pt0 -NORM=ext
-HEAPSIZE=500 -MTNLN=0 -MAXLEN=200000000
-USER=US10049750_QCGN_1_1_405@runat_23122005_113555_6388 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAREXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database : Published Applications AA Main:*

```

1: /cgn2_6/ptodata/1/pubpaa/us07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/1/pubpaa/us08_PUBCOMB.pcp.*
3: /cgn2_6/ptodata/1/pubpaa/us09_PUBCOMB.pcp.*
4: /cgn2_6/ptodata/1/pubpaa/us10a_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/1/pubpaa/us10b_PUBCOMB.pcp.*
6: /cgn2_6/ptodata/1/pubpaa/us11_PUBCOMB.pcp.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1519	95.5	306	4	US-10-238-075-1139	Sequence 1139, App	
2	870	54.7	298	4	US-10-369-493-11397	Sequence 11397, A	
3	870	54.7	298	4	US-10-369-493-14281	Sequence 14281, A	
4	870	54.7	298	4	US-10-369-493-14829	Sequence 14829, A	
5	870	54.7	298	4	US-10-369-493-15011	Sequence 15011, A	
6	797	50.1	304	4	US-10-369-493-7730	Sequence 7730, App	
7	550	34.6	308	3	US-09-815-245-11854	Sequence 11854, A	
8	546	34.3	222	5	US-10-781-581-208	Sequence 208, App	
9	546	34.3	322	5	US-10-745-237-304	Sequence 304, App	
10	523	32.9	343	4	US-10-369-493-5584	Sequence 5584, App	
11	507	31.9	309	4	US-10-369-493-12350	Sequence 12350, A	

ALIGNMENTS

RESULT 1

```

US-10-238-075-1139
; Sequence 1139, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of
; TITLE OF INVENTION: E.coli, and biological uses
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1139
; LENGTH: 306
; TYPE: prt
; ORGANISM: Escherichia coli
US-10-238-075-1139

```

Alignment Scores:		
Pred. No.:	1.2e-149	306
Score:	1519.00	294
Percent Similarity:	98.69%	8
Best Local Similarity:	96.08%	Conservative: 4
Query Match:	95.53%	Mismatches: 4
DB:	4	Indels: 0
		Gaps: 0
	Length:	
	Matches:	

US-10-049-750-11 (1-921) X US-10-238-075-1139 (1-306)

OV 1 ATGGATATCGCGGTATTGGCTCTAAACATGGTGACCTTATCACCTACCAACCAGATG 60

1 MetAspIleAlaValIleGlySerAsnMetValAspIleuIleThrTyrThrAsnGlnMet 20

```
QY 61 CCCAAAGAGGGGAAACTCTGGAAGCGCGCGCTTTAAATCGCGTCCGCGGAAAGGG 120
Db 21 ProLysGluGluThrLeuGluAlaProAlaPheIleGlyCysGlyGlyVal 40
QY 121 GCGAAACAGCCCTGCGCGCGCTTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTG 180
Db 41 AlaAsnGlnAlaValAlaAlaAlaLysLeuAsnSerLysValLeuMetLeuThrLysVal 60
QY 181 GCGCAGCATATTTTCCCGCACACACCATTCGAATCTCGAATCCTGGGCGGATCAATAG 240
Db 61 GlyAspAspIlePheAlaAspAsnThrIleArgAsnLeuGluSerTrpGlyIleAsnThr 80
QY 241 ACCTATGTAGAAAAGTACCCTGTACAGCAGCGCGGTAGCGCGATTTTCGTCAACGCC 300
Db 81 ThrTyrValGluLysValProCysThrSerSerGlyValAlaProIlePheValAsnAla 100
QY 301 AACTCCAGCAACAGCATTTCTGATCATCAAGGCGCTAAACAAGTTTCTCTCGCGGGAAGAT 360
Db 101 AsnSerSerAsnSerIleLeuIleLysGlyAlaAsnLysPheLeuSerProGluAsp 120
QY 361 ATCGATTCGCGCGCGGAAGATTTAAAAAATGCCAGCTTATTTCTTGCACCTGGGAAGTT 420
Db 121 IleAspArgAlaAlaGluAspLeuLysCysLysLeuIleValLeuGlnLeuGluVal 140
QY 421 CAGCTTGAACCGCTTATCACGCAATAGAAATTTGGCAAGAAACACACGGGATTGAAGTGT 480
Db 141 GlnLeuGluThrValTyrHisAlaIleGluPheGlyLysLysGlyIleGluValLeu 160
QY 481 TTAAACCCCTGCGCCAGCATTTACGGGAATTAGATATGCTTTATGCCCTGTAATTCGATTC 540
Db 161 LeuAsnProAlaProAlaLeuArgGluLeuAspMetSerTyrAlaCysLysCysAspPhe 180
QY 541 TTTGTACCTAATCAAAACGAGCTGGAATATTAACCGGTATGCCAGTGGATACCTATGAC 600
Db 181 PheIleProAsnGluThrGluLeuGluIleLeuThrGlyMetSerValAspThrTyrAsp 200
QY 601 CATATTCCGCGCAGCGGCGCTGCTGGTAGATAAAAGGCGTGAACAATATTATGTCAAC 660
Db 201 HisIleArgLeuAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnIleIleValThr 220
QY 661 ATGGCGCAGAAAGCGCGCTGTGCGATGACGCGTGACAGGAAGTCCATGTTCCGCGGTTT 720
Db 221 MetSerGluLysGlyAlaLeuTrpMetThrArgAspGlnGluValHisValProAlaPhe 240
QY 721 AGAGTCAACCTCTGTATACACGCGGCGCGGCGATCCCTTTATCGCTGTTTCGCGCAT 780
Db 241 LysValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheSerHis 260
QY 781 TACTAGCTCCAGAGCGGGATGTGGAAAGCCCATGAATAAGCCGCTCTCTTTGCGGCT 840
Db 261 TyrTyrValGlnSerGlyAspValGluAlaLeuLysLysAlaAlaLeuPheAlaAla 280
QY 841 TTCAGCGTCCACGGGAAAGCACCAATCCTTTATCCAGCATTGAGCAATTTAATGAG 900
Db 281 PheSerValThrGlyLysGlyThrGlnSerSerTyrProSerIleGluGlnPheAsnGlu 300
QY 901 TATCTTTCGTTGAACGA 918
Db 301 PheLeuThrLeuAsnGlu 306
```

RESULT 2

```
US-10-369-493-11397
; Sequence 11397, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
```

```
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11397
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-11397

Alignment Scores:
Pred. No.: 11e-81 Length: 298
Score: 870.00 Matches: 167
Percent Similarity: 76.61% Conservative: 59
Best Local Similarity: 56.61% Mismatches: 69
Query Match: 54.72% Indels: 0
DB: Gaps: 4

US-10-049-750-11 (1-921) x US-10-369-493-11397 (1-298)

QY 7 ATCGCGTTATTGGCTCTAACATGTGGACCTTATCACCTACACCAACAGATGCCAAA 66
Db 4 IleGlyValValGlySerAsnMetValAspLeuIleThrTyrValAspArgMetProGly 23
QY 67 GAAGGGCAAACTCTGGAAGCGCGCGTTTAAATCGGCTGCGCGGAAAAAGGGCGAAC 126
Db 24 ProGlyValThrLeuGluAlaProThrPheGluMetGlyCysGlyGlyLysGlyAlaAsn 43
QY 127 CAGGCGCTGCGCGCCCTAAGCTCAATCAAAAGTATTGATGTTGACCAAGTGGCGGAC 186
Db 44 GlnAlaValAlaAlaAlaArgLeuGlyAlaGluValMetMetValThrArgValGlyAsp 63
QY 187 GATATTTTTCGCGACACACCATTCCTTAATCTCGAATCTTGGGGGATCAATACGCGTAT 246
Db 64 AspValPheAlaAspAsnThrIleArgAsnLeuAlaSerPheGlyValAspThrArgHis 83
QY 247 GTAGAAAAAGTACCGGTGTACCAGCAGCGCGTAGCGCGATTTTCGTCAACGCCAACTCC 306
Db 84 ValValLysValSerGlyLysSerSerGlyValAlaProIlePheValGluGlnSerGly 103
QY 307 AGCAACAGCATTCGATCATCAAGCGCTAAACAGTTTCTCTCGCGGGAAGATATCGAT 366
Db 104 GluAsnSerIleLeuIleValLysGlyAlaAsnAlaAspLeuLeuProValGluValAsp 123
QY 367 CGCGCGCGGAGATTTAAAAATCCAGCTTATTTGTTCTGCACTGGAAGTTCAGCTT 426
Db 124 LysAlaAlaAlaAspLeuLysGluCysGlyLeuIleLeuMetGlnMetGluValProVal 143
QY 427 GAAACCGTTTATCACGCAATAGAAATTTGGCAAGAAACACGGGATTGAAGTGTATTAAAC 486
Db 144 GluThrValTyrHisThrIleGluPheAlaAlaGlnAsnGlyIleGluThrIleLeuAsn 163
QY 487 CCTGCCCGCAGCATTCACGGGAATTAGATATGCTTTATGCTGCTGTAATTCGGAATTTCTTTGTA 546
Db 164 ProAlaProAlaAlaAlaAsnLeuAspProGluArgIleArgGlnValThrPheLeuVal 183
QY 547 CCTAATGAAACCGAGCTGGAATATTAAACCGGTATGCCGATGCGATACCTATGACCATATT 606
Db 184 ProAsnGluSerGluLeuAlaLeuLeuSerGlyLeuProThrAspThrAspGluAspIle 203
QY 607 CGCGCAGCGCACGTTTCGCTGGTAGATAAAGGCTGCAACAATATTATTGTCACCATGGGC 666
Db 204 ValArgAlaAlaArgSerLeuIleAlaArgIleArgThrValIleValThrLeuGly 223
QY 667 GAGAAAGCGCGCTGTGGATGACGCGTGACAGGAAGTCCATGTTCCGCGGTTTAGATG 726
Db 224 AlaArgGlyAlaArgMetIleThrSerAspGluIleValAsnIleGluProValLysVal 243
QY 727 AACGCTGTTGATACACGCGCGCGGATGCTTTATCGGCTGTTTCGCGCATTACTAC 786
Db 244 ThrProArgAspThrThrGlyAlaGlyAspAlaPheIleGlySerPheAlaArgPheTyr 263
```

```
Qy 787 GTCAGAGCGGGGATGTGGAAGCGCCATGAAGAAAGCGCTCTCTTTGGCGCTTTTCAGC 846
Db 264 AlaGluThrArgGluValValSerSerLeuLysAlaSerLeuTyAlaAlaHisSer 283
Qy 847 GTCACCGGAAAGGACCCCAATCTCTTATCCAGCAATGAGCAA 891
Db 284 IleThrArgProGlyThrGlnLysAlaTyAlaSerIleAspGlu 298

RESULT 3
US-10-369-493-14281
; Sequence 14281, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 14281
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-14281

Alignment Scores:
Pred. No.: 1.1e-81 Length: 298
Score: 870.00 Matches: 167
Percent Similarity: 76.61% Conservatives: 59
Best Local Similarity: 56.61% Mismatches: 69
Query Match: 54.72% Indels: 0
DB: 4 Gaps: 0

US-10-049-750-11 (1-921) x US-10-369-493-14281 (1-298)
Qy 7 ATCGCGGTTATTGGCTTAACATGGTGGACCTTATCACTACCAACACAGATGCCAAA 66
Db 4 IleGlyValValGlySerAsnMetValAspLeuIleThrTyValAspArgMetProGly 23
Qy 67 GAAGGGAACCTCTGGAAGCGCGCTTTAAATCGCTGCGCGGAAAGGGCGAAC 126
Db 24 ProGlyGluThrLeuGluAlaProThrPheGluMetGlyCysGlyGlyLysGlyAlaAsn 43
Qy 127 CAGCGCGTGGCGCGCTAAAGCTCAATTTCAAAAGTATTGATTTGACCAAGTGGCGAC 186
Db 44 GlnAlaValAlaAlaAlaAArgLeuGlyAlaGluValMetMetValThrArgValGlyAsp 63
Qy 187 GATATTTTGGCGAACAACATTCGTAATCTCGAATCTCTGGGGGATCAATACGACGTAT 246
Db 64 AspValPheAlaAspAsnThrIleArgAsnLeuAlaSerPheGlyValAspThrArgHis 83
Qy 247 GTAGAAAGTACCGTCTACACAGCGCGTACGCGCGATTTTCGTCACGCACTCC 306
Db 84 ValValLysValSerGlyLysSerSerGlyValAlaProIlePheValGluGlnSerGly 103
Qy 307 AGCAACAGCATTTCTGATCATCAAGCGCTAAACAAAGTTTCTCGCGCGAAGATATCGAT 366
Db 104. GluAsnSerIleLeuIleValLysGlyAlaAsnAlaAspLeuLeuProValGluValAsp 123
Qy 367 CGCGCGCGGAAGATTTAAATAATGCCAGCTTATTTCTTTCGCACTGGAAGTTACGTT 426
Db 124 LysAlaAlaAspLeuLysGlyCysGlyLeuIleLeuMetGlnMetGluValProVal 143
Qy 427 GAAACGGTTTATTCACGCAATAGATTTGGCAAGAAACACGCGGATTAAGTTATTAAAC 486
Db 144 GluThrValTyHisThrIleGluPheAlaAlaGlnAsnGlyIleGluThrIleLeuAsn 163
```

```
Qy 487 CTGCGCGCAGCATTAACGGGAATTAGATATCTCTTATGCTGTAAATCGGATTTCTTTGTA 546
Db 164 ProAlaProAlaAlaAlaAsnLeuAspProGluArgIleArgGlnValThrPheLeuVal 183
Qy 547 CCTAATGAACCGAGCTGGAATATTAAACGGTATGCCAGTGGATCCTATGACCATATT 606
Db 184 ProAsnGluSerGluLeuAlaLeuLeuSerGlyLeuProThrAspThrAspGluAspIle 203
Qy 607 CGCGCAGCGCAGCTTCGCTGTAGATAAAGGCTGAACAATATTATTGTCCACCATGGC 666
Db 204 ValArgAlaAlaArgSerLeuIleAlaArgGlyIleArgThrValIleValThrLeuGly 223
Qy 667 GAGAAAGCGCGCTGTGATGACGCGTGACGACGAGAGTCCATGTTCCGCGGTTTAGAGTG 726
Db 224 AlaArgGlyAlaArgMetIleThrSerAspGluIleValAsnIleGluProValLysVal 243
Qy 727 AACGCTGTCATACCAAGCGCGCGATGCTCTTATCGCTGTTTTCGCGCATTTACTAC 786
Db 244 ThrProArgAspThrThrGlyAlaGlyAspAlaPheIleGlySerPheAlaArgPheTy 263
Qy 787 GTCAGAGCGGGATGTGGAAGCGCGCATGAAAGCCGCTCTCTTTGCGCTTTTCAGC 846
Db 264 AlaGluThrArgGluValValSerSerLeuLysLysAlaSerLeuTyAlaAlaHisSer 283
Qy 847 GTCACCGGAAAGCGCACCAATCCTCTTATCCAGCAATGAGCAA 891
Db 284 IleThrArgProGlyThrGlnLysAlaTyAlaSerIleAspGlu 298
```

RESULT 4

US-10-369-493-14829
; Sequence 14829, Application US/10369493
; Publication No. US20030233675A1

; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 14829
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-14829

Alignment Scores:
Pred. No.: 1.1e-81 Length: 298
Score: 870.00 Matches: 167
Percent Similarity: 76.61% Conservatives: 59
Best Local Similarity: 56.61% Mismatches: 69
Query Match: 54.72% Indels: 0
DB: 4 Gaps: 0

US-10-049-750-11 (1-921) x US-10-369-493-14829 (1-298)

```
Qy 7 ATCGCGGTTATTGGCTTAACATGGTGGACCTTATCACCTACCAACACAGATGCCAAA 66
Db 4 IleGlyValValGlySerAsnMetValAspLeuIleThrTyValAspArgMetProGly 23
Qy 67 GAAGGGAACCTCTGGAAGCGCGCGCTTTAAATCGCTGCGCGGAAAGGGCGAAC 126
Db 24 ProGlyGluThrLeuGluAlaProThrPheGluMetGlyCysGlyGlyLysGlyAlaAsn 43
Qy 127 CAGCGCGTGGCGCGCTAAAGCTCAATTTCAAAAGTATTGATTTGACCAAGTGGCGAC 186
```

```
Db      44  GlnAlaValAlaAlaAlaAaLeuGlyAlaGluValMetMetValThrArgValGlyAsp 63
Qy      187  GATATTTTCCGACAAACACCATTCGTAATCTCGAATCCTGGGGATCAATACAGCAT 246
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      64  AspValPheAlaAspAenThrIleArgAsnLeuAlaSerPheGlyValAspThrArgHis 83
Qy      247  GTAGAAAAGTACCGTGTACAGCAGCGGGGTAGCGCGGATTTTCGTCAACGCCAATCC 306
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      84  ValValysValSerGlyLysSerSerGlyValAlaProIlePheValGluGlnSerGly 103
Qy      307  AGCAACAGCATTTCTGATCATCAAGCGCTAAACAAGTTTCTCTCGCGGGAAGATATCAT 366
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      104  GluAsnSerIleLeuIleValysGlyAlaAsnAlaAspLeuProValGluValAsp 123
Qy      367  CGCGCGCGGAAGATTTAAAAAATGCCAGCTTATTCTTCTGCAACTGGAGTTACGCTT 426
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      124  LysAlaAlaAlaAspLeuLysGluCysGlyLeuLeuMetGlnMetGluValProVal 143
Qy      427  GAAACGGTTTATCAGCAATAGAAATTTGGCAAGAAACACCGGATTTGAAGTGTATTAAAC 486
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      144  GluThrValTyrHisThrIleGluPheAlaAlaGlnAsnGlyIleGluThrIleLeuAsn 163
Qy      487  CCTGCGCGCAGCATTTACGGGAATTAGATATGCTTATGCTGTAATGCGGATTTCTTTGTA 546
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      164  ProAlaProAlaAlaAlaAsnLeuAspProGluArgIleArgGlnValThrPheLeuVal 183
Qy      547  CCTAATGAACCGAGCTGGAAATATTAAACCGGTATGCCAGTGGATACCTATGACCATAT 606
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      184  ProAsnGluSerGluLeuAlaLeuLeuSerGlyLeuProThrAspThrAspGluAspIle 203
Qy      607  CGCGCAGCGCGCTCGCTGGTAGATAAAGGCGTGAACAATATTATTGTCAACATGGGC 666
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      204  ValArgAlaAlaArgSerLeuIleAlaArgIlyIleArgThrValIleValThrLeuGly 223
Qy      667  GAGAAAGCGCGCTGTGGATGACGTGACGAGAGTCCATGTTCCGGCGTTTAGAGTG 726
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      224  AlaArgGlyAlaArgMetIleThrSerAspGluIleValAsnIleGluProValLysVal 243
Qy      727  AACGCTGTTGATACCAAGCGCGCGGCGATGCTTTTATCGGCTGTTTCGGCGCATTTACTAC 786
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      244  ThrProArgAspThrThrGlyAlaGlyAspAlaPheIleGlySerPheAlaArgPheTyr 263
Qy      787  GTCAGAGCGGGATGTGGAGCGCCATGAANAAGCCCTCTCTTTGCGGCTTTTCAGC 846
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      264  AlaGluThrArgGluValValSerSerLeuLysAlaSerLeuLysAlaAlaHisSer 283
Qy      847  GTCACCGGAAAGCGCACCAATCTTATCCAGCATTTGAGCAA 891
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      284  IleThrArgProGlyThrGlnLysAlaTyrAlaSerIleAspGlu 298

RESULT 5
US-10-369-493-15011
; Sequence 15011, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15011
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-15011
```

```
Alignment Scores:
Pred. No.: 11e-81      Length: 298
Score: 870.00      Matches: 167
Percent Similarity: 76.61%      Conservative: 59
Best Local Similarity: 56.61%      Mismatches: 69
Query Match: 54.72%      Indels: 0
DB: 4      Gaps: 0

US-10-049-750-11 (1-921) x US-10-369-493-15011 (1-298)
```

```
Qy      7  ATCCGCGTTATTGGCTCTAAACATGGTGGAGCTTTATCACCTACCAACCAAGATGCCAAA 66
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      4  IleGlyValValGlySerAsnMetValAspLeuIleThrTyrValAspArgMetProGly 23
Qy      67  GAAGGGGAAACTCTGAAGCGCGGCTTTAAATCGGCTGCGCGGGAAGGGGCGAAC 126
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      24  ProGlyGluThrLeuGluAlaProThrPheGluMetGlyCysGlyGlyysGlyAlaAsn 43
Qy      127  CAGCGCGTGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGGCGAC 186
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      44  GlnAlaValAlaAlaAlaArgLeuGlyAlaGluValMetMetValThrArgValGlyAsp 63
Qy      187  GATATTTTCCGACAAACACCATTCGTAATCTCGAATCCTGGGGATCAATACAGCAT 246
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      64  AspValPheAlaAspAenThrIleArgAsnLeuAlaSerPheGlyValAspThrArgHis 83
Qy      247  GTAGAAAAGTACCGTGTACAGCAGCGGGGTAGCGCGGATTTTCGTCAACGCCAATCC 306
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      84  ValValysValSerGlyLysSerSerGlyValAlaProIlePheValGluGlnSerGly 103
Qy      307  AGCAACAGCATTTCTGATCATCAAGCGCTAAACAAGTTTCTCTCGCGGGAAGATATCAT 366
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      104  GluAsnSerIleLeuIleValysGlyAlaAsnAlaAspLeuProValGluValAsp 123
Qy      367  CGCGCGCGGAAGATTTAAAAAATGCCAGCTTATTGTTCTGCAACTGGAGTTACGCTT 426
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      124  LysAlaAlaAlaAspLeuLysGluCysGlyLeuLeuMetGlnMetGluValProVal 143
Qy      427  GAAACGGTTTATCAGCAATAGAAATTTGGCAAGAAACACCGGATTTGAAGTGTATTAAAC 486
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      144  GluThrValTyrHisThrIleGluPheAlaAlaGlnAsnGlyIleGluThrIleLeuAsn 163
Qy      487  CCTGCGCGCAGCATTTACGGGAATTAGATATGCTTATGCTGTAATGCGGATTTCTTTGTA 546
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      164  ProAlaProAlaAlaAlaAsnLeuAspProGluArgIleArgGlnValThrPheLeuVal 183
Qy      547  CCTAATGAACCGAGCTGGAAATATTAAACCGGTATGCCAGTGGATACCTATGACCATAT 606
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      184  ProAsnGluSerGluLeuAlaLeuLeuSerGlyLeuProThrAspThrAspGluAspIle 203
Qy      607  CGCGCAGCGCGCTCGCTGGTAGATAAAGGCGTGAACAATATTATTGTCAACATGGGC 666
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      204  ValArgAlaAlaArgSerLeuIleAlaArgIlyIleArgThrValIleValThrLeuGly 223
Qy      667  GAGAAAGCGCGCTGTGGATGACGTGACGAGAGTCCATGTTCCGGCGTTTAGAGTG 726
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      224  AlaArgGlyAlaArgMetIleThrSerAspGluIleValAsnIleGluProValLysVal 243
Qy      727  AACGCTGTTGATACCAAGCGCGCGGCGATGCTTTTATCGGCTGTTTCGGCGCATTTACTAC 786
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      244  ThrProArgAspThrThrGlyAlaGlyAspAlaPheIleGlySerPheAlaArgPheTyr 263
Qy      787  GTCAGAGCGGGATGTGGAGCGCCATGAANAAGCCCTCTCTTTGCGGCTTTTCAGC 846
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      264  AlaGluThrArgGluValValSerSerLeuLysAlaSerLeuLysAlaAlaHisSer 283
Qy      847  GTCACCGGAAAGCGCACCAATCTTATCCAGCATTTGAGCAA 891
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      284  IleThrArgProGlyThrGlnLysAlaTyrAlaSerIleAspGlu 298

RESULT 6
US-10-369-493-7730
```

; Sequence 7730, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 7730
; LENGTH: 304
; TYPE: PRT
; ORGANISM: Rhodobacter sphaeroides
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(304)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-7730

Alignment Scores:
Pred. No.: 4,896-74 Length: 304
Score: 797.00 Matches: 159
Percent Similarity: 70.65% Conservative: 46
Best Local Similarity: 54.83% Mismatches: 85
Query Match: 50.13% Indels: 0
DB: 4 Gaps: 0

US-10-049-750-11 (1-921) x US-10-369-493-7730 (1-304)

QY 7 ATCCGGTATTGGCTTAACATGCTGACCTTATCACTACCAACAGATGCCCAA 66
DB 15 llaAlaValValGlySerAsnMetValAspLeuValThrIleThrArgMetProAla 34
QY 67 GAAGGGAACTCTGGAAGCGCGGCTTTAAATCGGCTGGCGGGAAGGCGGCAAC 126
DB 35 ProGlyGluThrIleGluAlaProAspPheGluIleGlyCysGlyGlyGlyGlyAlaAsn 54
QY 127 CAGCGCTGGCGCGCTCAAGCTCAATCAAAAGATTGATCTGCACCAAGTGGCGCAC 186
DB 55 GlnAlaValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 74
QY 187 GATATTTTTCGCGACACACCACTTCGTAATCTCGAATCTCGGGGATCAATACGCGTAT 246
DB 75 AspIlePheGlyGluAsnThrArgAsnLeuAlaGlnHisGlyIleAspIleArgHis 94
QY 247 GTAGAAAAGTACCGTGTACAGCGCGGTAGCGCGATTTTGTCAACGCCCACTCC 306
DB 95 ValGluThrValAlaGlySerSerGlyValAlaProIlePheValGluAlaSerGly 114
QY 307 ACCAAGATTTCTGATCATCAAGCGCTACAGTTTCTCTCCGCGAAGATATCGAT 366
DB 115 GluAsnSerIleLeuIleValGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 134
QY 367 CCGCGCGCGGAGAGATTAAATAAATGCGAGCTTATTTGTTCTGCAACTGGAAGTTCAGCTT 426
DB 135 ThrAlaGluGluThrLeuArgAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 154
QY 427 GAAACGGTTTATCAGCAATGAATTTTGGCAGAAACACGGAATGAAGTGTATTAAAC 486
DB 155 GluThrValIleHisThrValArgAlaAlaGluThrGlyValArgThrIleLeuAsn 174
QY 487 CTGCGCGCAGCATTCAGGAATAGATATGTTCTTATGCTGTAAATGCGATTCTTTGTA 546
DB 175 ProAlaProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 194
QY 547 CCTAATGAAACCGAGCTGGAATATTAAACCGGTATGCCAGTGGATACCTATGACCATATT 606

DB 195 ProAsnGluSerGluLeuAlaLeuIleSerGlyLeuProThrGlySerGluGluIle 214
QY 607 CGCGCAGCGGCACGTTTCGCTGTAGATAAAGGGCTGAACAATATTATTGTCCATGGC 666
DB 215 AlaAlaAlaAlaLeuSerLeuIleGlyThrValIleValThrLeuGly 234
QY 667 GAGAAAGCGCGTGTGGATGACGCGTGACCAAGTCCATGTTCCGGCGTTTAGAGTG 726
DB 235 GlyArgGlyAlaArgLeuValThrArgAlaGlyValValProIleAlaProValArgVal 254
QY 727 AACGCTGTGATACCGCGCGCGGATGCGCTTTATCGGCTGTTTCGCGCATTCATAC 786
DB 255 ThrProValAspThrThrGlyAlaGlyAspAlaPheIleGlyAlaPheAlaHisPheLeu 274
QY 787 GTCCAGAGCGGGATGTGGAAGCGCCATGAAAGAAAGCGCTCTCTTTGCGCTTTCAGC 846
DB 275 AlaAlaThrGlyGluValGluGlyAlaLeuAlaHisAlaAlaArgTyrAlaAlaHisSer 294
QY 847 GTACCGGGAAGACCGCCCATCTCTTAT 876
DB 295 ValThrGlyArgGlyThrGlnLysSerTyr 304

RESULT 7

US-09-815-242-11854
; Sequence 11854, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Truwick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11854
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11854

Alignment Scores:
Pred. No.: 3,628-48 Length: 308
Score: 550.00 Matches: 126
Percent Similarity: 59.33% Conservative: 52
Best Local Similarity: 42.00% Mismatches: 122
Query Match: 34.59% Indels: 0
DB: 3 Gaps: 0

US-10-049-750-11 (1-921) x US-09-815-242-11854 (1-308)

```
QY 7 ATCCGGTTATTGGCTTAACATGCTGGACCTTATCACTACCAACACAGATGCCAAA 66
Db 7 ATCCGGTTATTGGCTTAAACATGCTGGACCTTATCACTACCAACACAGATGCCAAA 66
QY 5 ValLeuValValGlySerLeuMetAspLeuValValArgAlaProArgLeuProArg 24
Db 5 ValLeuValValGlySerLeuMetAspLeuValValArgAlaProArgLeuProArg 24
QY 67 GAAGGGGAAACTCTGGAAGCGCCGCGTTTAAATCGGCTCGCGGAAAGGGCGAAC 126
Db 67 GAAGGGGAAACTCTGGAAGCGCCGCGTTTAAATCGGCTCGCGGAAAGGGCGAAC 126
QY 25 GlyGlyGluThrLeuAlaGlyGlnSerPheThrValProGlyGlyLeuAlaAsn 44
Db 25 GlyGlyGluThrLeuAlaGlyGlnSerPheThrValProGlyGlyLeuAlaAsn 44
QY 127 CAGCGCGTGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTCACCAAGTGGCGAC 186
Db 127 CAGCGCGTGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTCACCAAGTGGCGAC 186
QY 45 GlnAlaValAlaAlaArgLeuGlyAlaGluValAlaMetIleGlyCysLeuGlyAsp 64
Db 45 GlnAlaValAlaAlaArgLeuGlyAlaGluValAlaMetIleGlyCysLeuGlyAsp 64
QY 187 GATATTTTTCGCGACCAACACCATTCGTAATCTCGAATCTCGGGGATCAATACGAGTAT 246
Db 187 GATATTTTTCGCGACCAACACCATTCGTAATCTCGAATCTCGGGGATCAATACGAGTAT 246
QY 65 AspAlaIleValGlyAspGlnLeuThrArgAlaLeuGlnAlaGluGlyIleAspCysGlnGly 84
Db 65 AspAlaIleValGlyAspGlnLeuThrArgAlaLeuGlnAlaGluGlyIleAspCysGlnGly 84
QY 247 GTAGAAAAGTACCGTGTACCAAGCGCGCTAGCGCGGATTTTCGTCACACGCCAATCC 306
Db 247 GTAGAAAAGTACCGTGTACCAAGCGCGCTAGCGCGGATTTTCGTCACACGCCAATCC 306
QY 85 ValGluArgValAlaGlyGluSerGlyValAlaLeuIleValValAspAspSer 104
Db 85 ValGluArgValAlaGlyGluSerGlyValAlaLeuIleValValAspAspSer 104
QY 307 AGCAACAGCATCTCATCAACAGCGCTAACAAGTTTCTCTCGCGGAAGATATCGAT 366
Db 307 AGCAACAGCATCTCATCAACAGCGCTAACAAGTTTCTCTCGCGGAAGATATCGAT 366
QY 105 GlnAsnAlaIleValIleValAlaGlyGlyAsnGlyHisLeuSerProAlaValAla 124
Db 105 GlnAsnAlaIleValIleValAlaGlyGlyAsnGlyHisLeuSerProAlaValAla 124
QY 367 CGCGCGCGGAGATTTAAATAATGCGAGCTTATTGTTCTGCAACTGCAAGTTTCAGCTT 426
Db 367 CGCGCGCGGAGATTTAAATAATGCGAGCTTATTGTTCTGCAACTGCAAGTTTCAGCTT 426
QY 125 ArgHisGluHisLeuLeuGluGlnAlaGlnValValValCysGlnLeuGluSerProLeu 144
Db 125 ArgHisGluHisLeuLeuGluGlnAlaGlnValValValCysGlnLeuGluSerProLeu 144
QY 427 GAAACGGTTTATCAGCAATAGAATTTGCGAAGAAACACGGGATTTGAAGTGTATTAAAC 486
Db 427 GAAACGGTTTATCAGCAATAGAATTTGCGAAGAAACACGGGATTTGAAGTGTATTAAAC 486
QY 145 GluThrValGlyHisValLeuArgAlaHisAlaLeuGlyLysThrValIleLeuAsn 164
Db 145 GluThrValGlyHisValLeuArgAlaHisAlaLeuGlyLysThrValIleLeuAsn 164
QY 487 CTGCGCGCAGCATTTACGGGAATTAGATATGCTTATGCTGTAAATGCGATTTCTTTGTA 546
Db 487 CTGCGCGCAGCATTTACGGGAATTAGATATGCTTATGCTGTAAATGCGATTTCTTTGTA 546
QY 165 ProAlaProAlaThrArgAspValProAlaGluTrpLeuProLeuValAspTyrLeuVal 184
Db 165 ProAlaProAlaThrArgAspValProAlaGluTrpLeuProLeuValAspTyrLeuVal 184
QY 547 CCTAATGAACCGAGCTGGAATATTAACCGTATGCGAGTATGCGAGTATGACCATATT 606
Db 547 CCTAATGAACCGAGCTGGAATATTAACCGTATGCGAGTATGCGAGTATGACCATATT 606
QY 185 ProAsnGluThrGluSerGlnLeuLeuCysArgLeuProValAspSerLeuGluSerAla 204
Db 185 ProAsnGluThrGluSerGlnLeuLeuCysArgLeuProValAspSerLeuGluSerAla 204
QY 607 CGCGCAGCGGCGATTCGCTGTAGATAAAGGGCTGAAACAATATTATTGTCCACCATGGC 666
Db 607 CGCGCAGCGGCGATTCGCTGTAGATAAAGGGCTGAAACAATATTATTGTCCACCATGGC 666
QY 205 GlyArgAlaAlaGluArgLeuGluMetGlyAlaGlyArgValIleValThrLeuGly 224
Db 205 GlyArgAlaAlaGluArgLeuGluMetGlyAlaGlyArgValIleValThrLeuGly 224
QY 667 GAGAAAGCGCGCTGTGATGACCGCTGACCGAAGTCCATGTTCCGCGGTTTGAAGTG 726
Db 667 GAGAAAGCGCGCTGTGATGACCGCTGACCGAAGTCCATGTTCCGCGGTTTGAAGTG 726
QY 225 AlaGlnGlyAlaLeuValGlyGluGlyArgValGluHisPheProValAlaArgVal 244
Db 225 AlaGlnGlyAlaLeuValGlyGluGlyArgValGluHisPheProValAlaArgVal 244
QY 727 AACGCTGTGTATACCAAGCGCGCGGATGCTTTATCGGCTGTTTCGCGCATTTACTAC 786
Db 727 AACGCTGTGTATACCAAGCGCGCGGATGCTTTATCGGCTGTTTCGCGCATTTACTAC 786
QY 245 LysAlaLeuAspThrThrAlaAlaGlyAspThrPheValGlyGlyPheAlaAlaAlaLeu 264
Db 245 LysAlaLeuAspThrThrAlaAlaGlyAspThrPheValGlyGlyPheAlaAlaAlaLeu 264
QY 787 GTCCAGAGCGGGGTGTGGAAGCGCCCATGAAAGGCGCTCTCTTTGCGCGTTTCAGC 846
Db 787 GTCCAGAGCGGGGTGTGGAAGCGCCCATGAAAGGCGCTCTCTTTGCGCGTTTCAGC 846
QY 265 AlaArgGlyLeuAspGluAlaAlaAlaIleArgPheGlyGlnAlaAlaAlaIleSer 284
Db 265 AlaArgGlyLeuAspGluAlaAlaAlaIleArgPheGlyGlnAlaAlaAlaIleSer 284
QY 847 GTCACCGGGAAGGACCAATCTCTTATTCACAGCATTTGAAGTATTTATGAGTATCTT 906
Db 847 GTCACCGGGAAGGACCAATCTCTTATTCACAGCATTTGAAGTATTTATGAGTATCTT 906
QY 285 ValThrArgLeuGlyAlaGlnThrSerIleProSerArgGluGluValGluArgAlaLeu 304
Db 285 ValThrArgLeuGlyAlaGlnThrSerIleProSerArgGluGluValGluArgAlaLeu 304
```

RESULT 8

```
US-10-781-581-208
; Sequence 208, Application US/10781581
; Publication No. US20050019746A1
; GENERAL INFORMATION:
; APPLICANT: BiRx Therapeutics Ltd.
; APPLICANT: Seery, Liam
; APPLICANT: Hayes, Ian
; APPLICANT: Murphy, Finbarr
; TITLE OF INVENTION: Apoptosis-Related Kinase/GPCRs
; FILE REFERENCE: 8912/2015
; CURRENT APPLICATION NUMBER: US/10/781,581
; CURRENT FILING DATE: 2004-02-18
```

```
; PRIOR APPLICATION NUMBER: US 10/764,238
; PRIOR FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US 60/457,533
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: UK 0301566.6
; PRIOR FILING DATE: 2003-01-23
; NUMBER OF SEQ ID NOS: 226
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 208
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-781-581-208

Alignment Scores:
Pred. No.: 9,67e-48 Length: 322
Score: 546.00 Matches: 117
Percent Similarity: 58.98% Conservative: 57
Best Local Similarity: 39.66% Mismatches: 117
Query Match: 34.34% Indels: 4
DB: 5 Gaps: 2
```

US-10-049-750-11 (1-921) x US-10-781-581-208 (1-322)

```
QY 7 ATCCGGTTATTGGCTTAAACATGCTGGACCTTATCACTACCAACACAGATGCCAAA 66
Db 7 ATCCGGTTATTGGCTTAAACATGCTGGACCTTATCACTACCAACACAGATGCCAAA 66
QY 18 ValValValValGlySerCysMetThrAspLeuValSerLeuThrSerArgLeuProLys 37
Db 18 ValValValValGlySerCysMetThrAspLeuValSerLeuThrSerArgLeuProLys 37
QY 67 GAAGGGGAAACTCTGGAAGCGCGCGTTTAAATTCGGCTCGCGGAAAGGGCGAAC 126
Db 67 GAAGGGGAAACTCTGGAAGCGCGCGTTTAAATTCGGCTCGCGGAAAGGGCGAAC 126
QY 38 ThrGlyGluThrIleHisGlyHisLysPhePheIleGlyPheGlyGlyAlaAsn 57
Db 38 ThrGlyGluThrIleHisGlyHisLysPhePheIleGlyPheGlyGlyAlaAsn 57
QY 127 CAGCGCGTGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGAC 186
Db 127 CAGCGCGTGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGAC 186
QY 58 GlnCysValGlnAlaAlaArgLeuGlyAlaMetThrSerMetValCysLysValGlyLys 77
Db 58 GlnCysValGlnAlaAlaArgLeuGlyAlaMetThrSerMetValCysLysValGlyLys 77
QY 187 GATATTTTTCGCGACCAACACCATTCGTAATCTCGAATCTCGGGGATCAATACGAGTAT 246
Db 187 GATATTTTTCGCGACCAACACCATTCGTAATCTCGAATCTCGGGGATCAATACGAGTAT 246
QY 78 AspSerPheGlyAsnAspTyrIleGluAsnLeuLysGlnAsnAspIleSerThrGluPhe 97
Db 78 AspSerPheGlyAsnAspTyrIleGluAsnLeuLysGlnAsnAspIleSerThrGluPhe 97
QY 247 GTAGAAAAGTACCGTGTACCAAGCGCGCTAGCGCGATTTTCGTCACACGCCAATCC 306
Db 247 GTAGAAAAGTACCGTGTACCAAGCGCGCTAGCGCGATTTTCGTCACACGCCAATCC 306
QY 98 ThrTyrGlnThrLysAspAlaAlaThrGlyThrAlaSerIleIleValAsnGluGly 117
Db 98 ThrTyrGlnThrLysAspAlaAlaThrGlyThrAlaSerIleIleValAsnGluGly 117
QY 307 AGCAACAGCATCTCATCAACAGCGCTAACAAGTTCCTCTCGCGGAGATATCGAT 366
Db 307 AGCAACAGCATCTCATCAACAGCGCTAACAAGTTCCTCTCGCGGAGATATCGAT 366
QY 118 GlnAsnIleIleValIleValAlaGlyAlaAsnLeuLeuLeuAsnThrGluAspLeuArg 137
Db 118 GlnAsnIleIleValIleValAlaGlyAlaAsnLeuLeuLeuAsnThrGluAspLeuArg 137
QY 367 CGCGCGCGGAGATTTAAATAATGCGAGCTTATTGTTCTGCAACTGGAAGTTCAGCTT 426
Db 367 CGCGCGCGGAGATTTAAATAATGCGAGCTTATTGTTCTGCAACTGGAAGTTCAGCTT 426
QY 138 AlaAlaAlaAsnValIleSerArgAlaLysValMetValCysGlnLeuGluIleThrPro 157
Db 138 AlaAlaAlaAsnValIleSerArgAlaLysValMetValCysGlnLeuGluIleThrPro 157
QY 427 GAAACGGTTTATCAGCAATAGAATTTGCGAAGAAACACCGGATTTGAAGTGTATTAAAC 486
Db 427 GAAACGGTTTATCAGCAATAGAATTTGCGAAGAAACACCGGATTTGAAGTGTATTAAAC 486
QY 158 AlaThrSerLeuGluAlaLeuThrMetAlaArgArgSerGlyValLysThrLeuPheAsn 177
Db 158 AlaThrSerLeuGluAlaLeuThrMetAlaArgArgSerGlyValLysThrLeuPheAsn 177
QY 487 CTGCGCGCAGCATTTACGGGAATTAGATATGCTTATGCTGTAAATGCGATTTCTTTGTA 546
Db 487 CTGCGCGCAGCATTTACGGGAATTAGATATGCTTATGCTGTAAATGCGATTTCTTTGTA 546
QY 178 ProAlaProAlaIleAlaAspLeuAspProGlnPheTyrThrLeuSerAspValPheCys 197
Db 178 ProAlaProAlaIleAlaAspLeuAspProGlnPheTyrThrLeuSerAspValPheCys 197
QY 547 CCTAATGAACCGAGCTGGAATATTAACCGGTATGCGAGTATGCGATCATCATGACCATATT 606
Db 547 CCTAATGAACCGAGCTGGAATATTAACCGGTATGCGAGTATGCGATCATCATGACCATATT 606
QY 198 CysAsnGluSerGluAlaGluIleLeuThrGlyLeuThrValGlySerAlaAlaAspAla 217
Db 198 CysAsnGluSerGluAlaGluIleLeuThrGlyLeuThrValGlySerAlaAlaAspAla 217
QY 607 CGCGCAGCGGCGATTCGCTGTAGATAAAGGGCTGAAACAATATTATTGTCCACCATGGC 666
Db 607 CGCGCAGCGGCGATTCGCTGTAGATAAAGGGCTGAAACAATATTATTGTCCACCATGGC 666
QY 218 GlyGluAlaAlaLeuValLeuLysArgGlyCysGlnValValIleIleThrLeuGly 237
Db 218 GlyGluAlaAlaLeuValLeuLysArgGlyCysGlnValValIleIleThrLeuGly 237
QY 667 GAGAAAGC-----GCCTGTGGATGACGCGTACAGGAGTCCATGTCCTTCGCGCTT 720
Db 667 GAGAAAGC-----GCCTGTGGATGACGCGTACAGGAGTCCATGTCCTTCGCGCTT 720
QY 238 AlaGluGluCysValValLeuSerGlnThrGluProGluProLysHisIleThrGlu 257
Db 238 AlaGluGluCysValValLeuSerGlnThrGluProGluProLysHisIleThrGlu 257
QY 721 AGAGTGAACGCTGTGTATACCAAGCGCGCGATGCTTTTATCGGCTGTTTCGCGCAT 780
Db 721 AGAGTGAACGCTGTGTATACCAAGCGCGCGATGCTTTTATCGGCTGTTTCGCGCAT 780
```


258	LysValIysAlaValAspThrThrGlyAlaGlyAspSerPheValGlyAlaLeuAlaPhe	277
781	TAC-----TAGCTCCAGACGGGGATCTGGGAAGCGGCATCAAAAAGACCGCTCTCTTT	834
278	TyrLeuAlaIrrTyrProAsnLeuSerLeuGluAspMetLeuAsnArgSerAsnPheIle	297
835	GCCGCTTTTCAGCGTCACCGGGGAAGCACCCCAATCCTCTTTATPCCA	879
298	AlaAlaValSerValGlnAlaAlaGlyThrGlnSerSerTyrPro	312

```

RESULT 9
US-10-745-237-304
; Sequence 304, Application US/10745237
; Publication No. US20050227301A1
; GENERAL INFORMATION:
; APPLICANT: Cyclacel Limited
; APPLICANT: Glover, David
; APPLICANT: Bell, Graham
; APPLICANT: Frenz, Lisa
; APPLICANT: Midgley, Carol
; TITLE OF INVENTION: Cell Cycle Progression Proteins
; FILE REFERENCE: P015819WO CYK
; CURRENT APPLICATION NUMBER: US/10/745,237
; CURRENT FILING DATE: 2003-12-23
; PRIOR APPLICATION NUMBER: US 60/439,123
; PRIOR FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: US 60/468,402
; PRIOR FILING DATE: 2003-05-06
; NUMBER OF SEQ ID NOS: 600
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 304
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Q9H477
US-10-745-237-304

```

367 CGCGCGCGGAAGATTTAAAAAATGCCAGCTTATTGTTCTGCAACTCGAAGCTTCAGCTT 426

138 AlaAlaAsnValIIeSerArgAlaValMetValCysGlnLeuGluIleThrPro 157

427 GAAACGGTTTTACACGCAATAGAATTTGGCAAGAAACACGCGATTGAAGTGTATTAAAC 486

158 AlaThrSerLeuGluAlaLeuThrMetAlaArgSerGlyValIleThrLeuPheAsn 177

487 CTCTGCGCCAGCATTTACGGGAATAGATATGCTCTATGCTGTAAATGCCATTCTTTGTA 546

178 ProAlaProAlaIIeAlaAspLeuAspProGlnPheThrLeuSerAspValPheCys 197

547 CCTAATGAAACCGAGCTGGAATATTAAACGGTATGCCGATGATACCTTACCAATATT 606

198 CysAsnGluSerGluAlaGluIleLeuThrGlyLeuThrValGlySerAlaAlaAspAla 217

607 CGCGCAGCGGCAGCTTGCCTGCTAGATAAAGCGGTGAACAATATTATGTCACCATGGGC 666

218 GlyGluAlaAlaLeuValLeuLeuAspGlyCysGlnValIleThrLeuGly 237

667 GAGAAAGGC-----GGCTCTGGATGACGGGTGACACGAGGAAGTCATGTTCCGGCGTTT 720

238 AlaGluGlyCysValValLeuSerGlnThrGluProGluProLysHisIleProThrGlu 257

721 AGATGAACCGCTGTTGTATACACGCGCGCGCGGCGATGCTTTATCGGCTGTTTCGGCGCAT 780

258 LysValIysAlaValAspThrThrGlyAlaGlyAspSerPheValGlyAlaLeuAlaPhe 277

781 TAC-----TAGTCCAGAGCGGGGATGTGGAAGCGGCCATCAAAAAAGCCGCTCTTT 834

278 TyrLeuAlaIleThrProAsnLeuSerLeuGluAspMetLeuAsnArgSerAsnPheIle 297

835 GCCGCTTTTCAGCGTCACCGGAAAGGCCACCCCAATCCTCTTATCCA 879

298 AlaAlaValSerValGlnAlaAlaGlyThrGlnSerSerTyrPro 312

RESULT 10
US-10-369-493-5584
; Sequence 5584, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5584
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5584

Alignment Scores:

Pred. No.:	2,54e-45	Length:	343
Score:	523.00	Matches:	114
Percent Similarity:	56.54%	Conservative:	59
Best Local Similarity:	37.25%	Mismatches:	127
Query Match:	32.89%	Indels:	6
DB:	4	Gaps:	2

[illegible]

```
Qy 67 GAAGGGAAACTCTGGAAGCGCGCGTCTTAAATCGGCTCGCGGAAAGGGCGAAC 126
    |||||:::
Db 57 ProGlyGluSerValArgGlyHisAsnPhelysLeuGlySerGlyGlyValAsn 76
    |||||:::
Qy 127 CAGCGCGTGGCGCGCTAAGCTCAATCAAAAGTATTGATTTGACCAAGTGGCGAC 186
    |||||:::
Db 77 GlnAlaValAlaAlaArgLeuGlyIleAsnValSerMetIleGlyMetValGlyGlu 96
    |||||:::
Qy 187 GATATTTTTCGGACACACCATTCGTAATCTCGAATCTCGGATCTGGGATCAATACACGTAT 246
    |||||:::
Db 97 AspMetPheGlyAspSerAsnIleGlyAspLeuSerSerAsnGlyValAspThrSerCys 116
    |||||:::
Qy 247 GTAGAAAAGTACCGTGTACACGACGCGGTAGCGCGATTTTCGTCAACGCCCACTCC 306
    |||||:::
Db 117 ValGlyArgThrLysLysThrHisThrAlaThrAlaThrIleThrValAsnLysGluGly 136
    |||||:::
Qy 307 AGCAACAGCATTCATCATCAAAAGCGCTAACAGTTTCTCTCGCGGAAGATATCGAT 366
    |||||:::
Db 137 GluAsnAsnIleValThrLeuGlyAlaAsnLeuGluMetSerProGluIleAlaAsp 156
    |||||:::
Qy 367 CGCGCGCGGAAGATTAAATAATGCCAGCTTATTGTTCTGCACTGGAAGTTCAGCTT 426
    |||||:::
Db 157 AlaAsnSerSerLysIleAlaGlySerLysMetValIleCysGlnGlyGluIleAspGlu 176
    |||||:::
Qy 427 GAAACGGTTTATCAGCAATAGATTTCGGCAAGAAACACGGGATTGAAGTGTATTAAAC 486
    |||||:::
Db 177 LysGlyAsnArgArgAlaPheGluIleAlaArgSerHisGlyValThrPheLeuAsn 196
    |||||:::
Qy 487 CTGCGCGCAGCATTTACGGGAATTAGATATGCTTATGCTGCTGTAATGCGATTTCTTTGTA 546
    |||||:::
Db 197 ProAlaProGlyAspProAsnMetAspLysThrIleLeuGluLeuValAspIleIleCys 216
    |||||:::
Qy 547 CCTAATGAAACCGAGCTGGAATATTAAACGGTATGACGATGCGATGATGACCATATT 606
    |||||:::
Db 217 ThrAsnGluAsnGluAlaGluPheIleThrGlyIleAlaGlnAsnAsnValLysAspAla 236
    |||||:::
Qy 607 CGCGCAGCGGACGTCGCTGCTAGATAAAGCGGTGAACAATATTATTGTCACCATGGC 666
    |||||:::
Db 237 GluLysAlaAlaArgAlaMetLeuLysMetGlyProGlnHisAlaIleIleThrLeuGly 256
    |||||:::
Qy 667 GAGAAAGCGCGCTGTGGATGACCGCT-----GACCAGGAAGTCCATGTTCCGCGCTT 720
    |||||:::
Db 257 GlyLysGlyValLeuLeuAlaSerLysGlyValAspValGluHisThrAlaValIle 276
    |||||:::
Qy 721 AGATGAAACGCTGTTGATACACGCGCGCGGCGATGCTTTATCGGCTGTTTCGCGCAT 780
    |||||:::
Db 277 LysValAspAlaValAspThrThrGlyAlaGlyAspCysPheCysGlySerLeuAlaAla 296
    |||||:::
Qy 781 TACTACGTCCAGAGCGGGATGTGGAAGCCGCCATGAAAGCGCTCTCTTTGCGCGCT 840
    |||||:::
Db 297 HisLeuValAlaGlyHisProIleSerAlaSerIleArgSerAlaAlaAsnLeuAlaAla 316
    |||||:::
Qy 841 TTCACGCTCACCGGAAAGGACCAATCTCTTATCCAAACATT-----GAG 888
    |||||:::
Db 317 LeuSerValThrArgHisGlyThrGlnSerSerTyrTrpLysLeuAspGluIleArgGln 336
    |||||:::
Qy 889 CAATTTAATGATATCTT 906
    |||||:::
Db 337 GlnTyrProGluPheLeu 342
    |||||:::
```

RESULT 11

```
US-10-369-493-12350
; Sequence 12350, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
```

```
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12350
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Mesorhizobium loti
US-10-369-493-12350

Alignment Scores:
Pred. No.: 1.16e-43 Length: 309
Score: 507.00 Matches: 113
Percent Similarity: 57.10% Conservative: 60
Best Local Similarity: 37.29% Mismatches: 126
Query Match: 31.89% Indels: 4
DB: 4 Gaps: 4

US-10-049-750-11 (1-921) x US-10-369-493-12350 (1-309)
```

```
Qy 7 ATCCGCGTATTGCTCTTAACATGCTGACCTTATCACTACACCAACAGATGCCCAA 66
    :::::
Db 7 ValValIleLeuGlyValPheValAlaAspThrAlaTyrArgAlaAspArgGlnProArg 26
    :::::
Qy 67 GAAGGGAAACTCTGGAAGCGCGCGTCTTAAATCGGCTCGCGGAAAGGGCGAAC 126
    |||||:::
Db 27 MetGlyGluThrIleLeuGlyAsnSerPheLysLeuGlyProGlyGlyLysGlySerAsn 46
    |||||:::
Qy 127 CAGCGCGTGGCGCGCTCAAGCTCAATCAAAAGTATTGATTTGACCAAGTGGCGAC 186
    |||||:::
Db 47 GlnAlaValAlaAlaGlyLysLeuGlyAlaAspThrThrPheLeuThrArgLeuGlyVal 66
    |||||:::
Qy 187 GATATTTTTCGGACACACCATTCGTAATCTCGAATCTCGGATGATGATGATGATGAT 246
    |||||:::
Db 67 AspAlaPheAlaAspMetAlaLysArgThrTrpGlnAspAlaGlyValLysSerAlaVal 86
    |||||:::
Qy 247 GTAGAAAAGTACCGTGTACACGACGCGGTAGCGCGATTTTCGTCAACGCCCACTCC 306
    |||||:::
Db 87 IleAsp---ThrProAspSerTyrThrGlyAlaAlaTyrIlePheValGluGluGlySer 105
    |||||:::
Qy 307 AGCAACAGCATTCATCATCAAA---GGCGCTAACAAAGTTTCTCTCGCGGAAGATATC 363
    |||||:::
Db 106 GlyAsnAsnAlaIleIleValSerProGlyAlaAlaMetLeuIleSerProThrAspIle 125
    |||||:::
Qy 364 GATCGCGCGCGGAAGATTAAATAATGCCAGCTTATTGTTCTGCACTCGAAGTTCAG 423
    |||||:::
Db 126 GluAlaAsnAlaAlaLeuIleArgGlyAlaGlyValPheValThrGlnLeuGluGlnPro 145
    |||||:::
Qy 424 CTTCAACCGGTTTATCAGCAATAGAAATTTGGCAAGAAACACGCGATTGAAGTGTATTA 483
    |||||:::
Db 146 IleAspAlaAlaMetArgAlaLeuGluIleAlaArgGlyAlaGlyValThrThrIleLeu 165
    |||||:::
Qy 484 AACCTCTCGCGCAGCATTTACGGGAATTAGATATGCTTATGCTGCTGTAATGCGATTCTTT 543
    |||||:::
Db 166 AsnProAlaProAlaAlaLysLeuProAspArgValTyrThrLeu---CysAspTyrLeu 184
    |||||:::
Qy 544 GTACCTAATGAACCGAGCTGGAATAATTAAACGGTATGCCAGTGATGATGATGATGATGAT 603
    |||||:::
Db 185 ThrProAsnGluThrGluThrGluGluLeuThrGlyLeuLysValSerSerValAspGlu 204
    |||||:::
Qy 604 ATTGCGCGCAGCGCACGTTTCGCTAGATAAAGCGGTGAACAATATTATTGTCACCATG 663
    |||||:::
Db 205 AlaArgThrAlaSerGlyLysLeuGluGlyValGlyThrValIleValThrLeu 224
    |||||:::
Qy 664 GCGGAGAAAGCGCGCTGTGGATCGCTGACCGGACCAAGTCCATGTTCCGCGCTTTAGA 723
    |||||:::
Db 225 GlyAspLysGlyAlaLeuLeuHisSerLysGlyArgSerGluHisValProAlaIleSer 244
    |||||:::
Qy 724 GTGAAC---GCTGTTGATACACGCGCGCGCGCATGCTTTATCGGCTGTTTCGCGCAT 780
    |||||:::
Db 245 AlaGlyProValValGluThrThrGlyAlaGlyAspAlaPheAsnGlyGlyPheAlaAla 264
    |||||:::
```



```
Db 4 ValValIleLeuGluGlyIleTyrValThrAspLeuThrPheArgAlaGlyArgMetProGln 23
Qy 67 GAAGGGAACTCTGGAAGCCCGCGCTTTAAATCGCTCGCGGGAAGGGCGGCAAC 126
Db 24 IleGlyGluThrIleAlaGlySerAlaPheAlaMetGlyProGlyGlySerAsn 43
Qy 127 CAGCGCGGCGCGCTGAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGAC 186
Db 44 GlnAlaValAlaAlaAlaArgAlaGlyAlaAspValPheCysThrArgIleGlyAsn 63
Qy 187 GATATTTTTCGCGACAAACACCATTCGTAATCTCGAATCCTCGGGGATCAATACGACGTAT 246
Db 64 AspAlaPheGlyAlaIleAlaArgAlaThrTrpAlaAlaGluGlyIle---ThrAlaArg 82
Qy 247 GTAGAAAAGTACCGTGTACGAGCGCGCTAGCGGATTTTCGTCAACGCCCACTCC 306
Db 83 AlaSerValIleGluGlyValSerThrGlyAlaAlaHisIlePheValAspAsnThr 102
Qy 307 AGC---AACACGATTCGATCATCAAGCGCGCTAACAGTTTCTCTCGCGGAAGATATC 363
Db 103 GlyMetAsnAlaIleIleValAlaSerGlyAlaAlaGlyThrMetGluProAlaAspVal 122
Qy 364 CATCGCGCGCGGAAGATTAAAAAATGCCAGCTTATTGTCTGCAACTCGAAGTTTCAG 423
Db 123 AspAlaIleGluAlaAspIleAlaAlaAlaArgValPheValThrGlnLeuGluGlnPro 142
Qy 424 CTTGAAACGGTTTATCAGCAATAGAAATTTGGCAAGAAACACGGGATTTGAAGTTTATTA 483
Db 143 LeuGluAlaAlaArgGlyLeuGluValAlaAlaArgHisGlyValIleThrValPhe 162
Qy 484 AACCTCGCGCAGCATTTACGGGAATTAGATATGCTTATGCTGTAATGCGATTTCTTT 543
Db 163 AsnProAlaProAlaLeu---ProLeuAspGlyIlePheProLeuCysAspIle 181
Qy 544 GTACTAATGAACCGAGCTGGAAATATTAACCGGTATGCCAGTGGATACCTATGACCAT 603
Db 182 ThrProAsnGluThrGluAlaThrAlaLeuThrGlyValProIleAlaAsnAlaAsp 201
Qy 604 ATTGCGCAGCGGCGCTGCTGCTGATAGTAAGGCTGAACATATTTATGTCACCATG 663
Db 202 AlaArgArgAlaAlaAspValLeuAlaAlaGlyValGlyValIleValThrLeu 221
Qy 664 GCGGAGAAAGCGCGCTGCGATGACGCGTACGAGGAGTCCATGTTCCGCGCTTAGA 723
Db 222 GlyGluGlyAlaLeuLeuHisSerAlaAlaGlnSerIleLeuValProAlaIleHis 241
Qy 724 GTG---AACCGCTGTTGATACCGCGCGCGGATGCTTTTATCGGCTGTTTCGCGCAT 780
Db 242 CysGlyArgValValGluThrAlaGlyAlaGlyAspGlyPheThrGlyGlyPheAlaAla 261
Qy 781 TACTACGTCCAGAGCGGGGTGTGAAGCGCGCATGAAAGGCGTCTCTTTGCGCGCT 840
Db 262 AlaLeuAlaArgGlyAspAlaMetThrAlaLeuArgPheGlyCysAlaLeuAlaGly 281
Qy 841 TTCAGCTGCACGGGAAGGACCACTCTTATCCAGCATTGACCAATTTAATGAG 900
Db 282 IleSerValThrArgProGlyThrAlaProSerMetProThrLeuAspGluValAsnGln 301
Qy 901 TATCTT 906
Db 302 ValLeu 303
```

RESULT 14

```
US-10-264-237-2526
; Sequence 2526, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA131P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; PRIORITY FILING DATE: 2002-10-04
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: Patent In Ver. 3.1
; SEQ ID NO 2526
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (213)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (215)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (217)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (282)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (285)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (296)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (297)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (299)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-237-2526
```

```
Alignment Scores:
Pred. No.: 4,85e-43 Length: 299
Score: 501.00 Matches: 104
Percent Similarity: 59.70% Conservative: 53
Best Local Similarity: 39.54% Mismatches: 104
Query Match: 31.51% Indels: 2
DB: 4 Gaps: 1
```

US-10-049-750-11 (1-921) x US-10-264-237-2526 (1-299)

```
Qy 7 ATCCGCGTATTGCTCTTAACATGGTGACCTTATCACCTACACCAACACGATGCCCAA 66
Db 18 ValValValValGlySerCysMetThrAspLeuValSerLeuThrSerArgLeuProlys 37
Qy 67 GAAGGGAACTCTGGAAGCCCGCGCTTTAAATCGCTCGCGGGAAGGGCGGCAAC 126
Db 38 ThrGlyGluThrIleHisGlyHisPhePheIleGlyPheGlyGlyAlaAsn 57
Qy 127 CAGCGCTGCGCGCTGAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGAC 186
Db 58 GlnCysValGlnAlaAlaArgLeuGlyAlaMetThrSerMetValCysLysValGlyLys 77
Qy 187 GATATTTTTCGCGACAAACACCATTCGTAATCTCGAATCCTCGGGGATCAATACGACGTAT 246
Db 78 AspSerPheGlyAsnAspTyrIleGluAsnLeuLysGlnAsnAspIleSerThrGluPhe 97
Qy 247 GTAGAAAAGTACCGTGTACGAGCGCGGTAGCGGATTTTCGTCAACGCCCACTCC 306
Db 98 ThrTyrGlnThrLysAspAlaAlaThrGlyThrAlaSerIleIleValAsnAsnGluGly 117
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 23, 2005, 22:44:48 ; Search time 7.5438 Seconds
(without alignments)
1741.646 Million cell updates/sec

Title: US-10-049-750-11
Perfect score: 1590
Sequence: 1 atggatcgcgggtattgg.....atcttcttgtaacgaataa 921

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Fgapop 10.0 , Fgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 54001 seqs, 7132810 residues

Total number of hits satisfying chosen parameters: 108002

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlp
-MODEL=frame+n2p.model -US10049750/runat_23122005_113555_6421/app_query.fasta_1.1742
-Q=/cgn2_1/USPTO_spool_p/US10049750/runat_23122005_113555_6421/app_query.fasta_1.1742
-DB=Published Applications AA New -QFMT=fastan -SUFFIX=n2p.rapbn -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10049750@cgn 1.1 @runat_23122005_113555_6421
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DETIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA New.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp.*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/2/pubpaa/PTC_NEW_PUB.pcp.*
5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp.*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/2/pubpaa/US6_NEW_PUB.pcp.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	458	28.8	302	US-10-793-626-3062	Sequence 3062, App
2	234	14.7	319	US-10-793-626-786	Sequence 786, App
3	167.5	10.5	238	US-10-485-517-177	Sequence 177, App
4	141.5	8.9	306	US-10-485-517-185	Sequence 185, App
5	118.5	7.5	346	US-10-467-657-6514	Sequence 6514, App
6	101	6.4	471	US-10-467-9628-10	Sequence 10, Appl
c	7	100.5	6.2	4384	Sequence 1120, App
8	91	5.7	187	US-10-793-626-448	Sequence 448, App
9	89	5.6	320	US-10-467-657-3254	Sequence 3254, App
10	87.5	5.5	445	US-10-858-730-209	Sequence 209, App

11	86	5.4	482	6	US-10-467-657-7660	Sequence 7660, Ap
12	85	5.3	186	6	US-10-793-626-2530	Sequence 2530, Ap
13	83.5	5.3	445	7	US-11-055-822-170	Sequence 170, App
14	83	5.2	244	6	US-10-467-657-3330	Sequence 3330, Ap
15	83	5.2	400	6	US-10-467-657-2702	Sequence 2702, Ap
16	82	5.2	317	6	US-10-485-517-196	Sequence 196, App
17	79.5	5.0	205	7	US-11-055-822-548	Sequence 548, App
18	79.5	5.0	205	7	US-11-055-822-1128	Sequence 1128, Ap
19	79.5	5.0	224	7	US-11-055-822-546	Sequence 546, App
20	79.5	5.0	224	7	US-11-055-822-1126	Sequence 1126, Ap
21	79.5	5.0	424	6	US-10-467-657-2678	Sequence 2678, Ap
22	79.5	5.0	525	7	US-11-074-176-146	Sequence 146, App
23	78.5	4.9	831	6	US-10-467-657-4486	Sequence 4486, Ap
24	78	4.9	233	6	US-10-793-626-416	Sequence 416, App
c	25	77.5	444	7	US-11-074-176-170	Sequence 170, App
26	77.5	4.9	445	6	US-10-873-528-30	Sequence 30, Appli
27	77.5	4.9	782	6	US-10-972-053-2	Sequence 2, Appli
28	77.5	4.9	782	6	US-10-972-053-8	Sequence 8, Appli
29	77	4.8	374	7	US-11-055-822-968	Sequence 968, App
c	30	77	1070	7	US-11-000-463-721	Sequence 721, App
31	77	4.7	1194	7	US-11-107-028-2	Sequence 2, Appli
32	76.5	4.8	798	7	US-11-107-028-2	Sequence 2, Appli
33	76.5	4.8	2214	7	US-11-080-991-94	Sequence 94, Appl
34	76	4.8	305	6	US-10-793-626-1362	Sequence 1362, Ap
35	76	4.8	400	6	US-10-467-657-7096	Sequence 7096, Ap
36	76	4.8	470	6	US-10-467-657-8420	Sequence 8420, Ap
37	76	4.8	625	6	US-10-793-626-2464	Sequence 2464, Ap
38	75.5	4.7	784	6	US-10-972-053-10	Sequence 10, Appl
39	75.5	4.7	792	6	US-10-972-053-12	Sequence 12, Appl
40	75	4.7	299	6	US-10-858-730-17	Sequence 17, Appl
41	75	4.7	343	6	US-10-467-657-4824	Sequence 4824, Ap
42	75	4.7	343	6	US-10-467-657-7256	Sequence 7256, Ap
43	75	4.7	348	6	US-10-467-657-4840	Sequence 4840, Ap
44	75	4.7	426	6	US-10-467-657-1548	Sequence 1548, Ap
45	75	4.7	436	6	US-10-467-657-7694	Sequence 7694, Ap

ALIGNMENTS

RESULT 1
US-10-793-626-3062
; Sequence 3062, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3062
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-3062

Alignment Scores:
Pred. No.: 5.09e-36 Length: 302
Score: 458.00 Matches: 108
Percent Similarity: 55.96% Conservative: 61
Best Local Similarity: 35.76% Mismatches: 129
Query Match: 28.81% Indels: 4
DB: 6 Gaps: 4

US-10-049-750-11 (1-921) x US-10-793-626-3062 (1-302)

QY 7 ATCGCGGTATTGGCTCTTAACATGGTGGACCTTACCTACCAACGATGCCAAA 66

Db 232 AlaThrIleTyrThrLysAspAspTyrArgIleHisHisGluGlyTyrGlnValGlnAla 251
 QY 733 GTTGATACACGCGCGCGGCGATCCCTTTATCGCTGTTCGCGCATTTACTACCTCCAG 792
 Db 252 IleAspThrThrGlyAlaGlyAspAlaPheIleGlyAlaIleTyrCysIleLeuGlu 271
 QY 793 AGC-----GGGATGTGAGCGCGCATGAAAGCGCTCTC----- 831
 Db 272 SerArgHisSerGluCysLysAspLeuPheLysGluLysGlyLysAspIleLeuAlaPhe 291
 QY 832 -----TTTGCGCGCTTTCAGCGTCACCGGAAAGGCACCAATCTCTTTATCCAAAGC 882
 Db 292 SerAsnArgValAlaAlaLeuThrThrLysHisGlyAlaIleGluSerLeuProThr 311
 QY 883 ATTGAGCAATTTAATGAGTAT 903
 Db 312 LysGluAspIleLysAspTyr 318
 RESULT 3
 US-10-485-517-177
 ; Sequence 177, Application US/10485517
 ; Publication No. US20050256299A1
 ; GENERAL INFORMATION:
 ; APPLICANT: University of Sheffield
 ; APPLICANT: Biosynexus Incorporated
 ; APPLICANT: Foster, Simon
 ; APPLICANT: Mond, James
 ; TITLE OF INVENTION: Antigenic Polypeptides
 ; FILE REFERENCE: P100629W0
 ; CURRENT APPLICATION NUMBER: US/10/485,517
 ; PRIOR FILING DATE: 2004-02-02
 ; PRIOR APPLICATION NUMBER: GB 0118825.9
 ; PRIOR FILING DATE: 2001-08-02
 ; PRIOR APPLICATION NUMBER: GB 0200349.9
 ; PRIOR FILING DATE: 2002-01-09
 ; NUMBER OF SEQ ID NOS: 424
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 177
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 US-10-485-517-177
 Alignment Scores:
 Pred. No.: 1,44e-08 Length: 238
 Score: 167.50 Matches: 52
 Percent Similarity: 48.00% Conservative: 56
 Best Local Similarity: 23.11% Mismatches: 92
 Query Match: 10.53% Indels: 25
 DB: 6 Gaps: 7
 US-10-049-750-11 (1-921) x US-10-485-517-177 (1-238)
 QY 289 TTCGTCAACGCACTCC---AGCAACAGCATTTGATCATCAAGCGCCTAAC----- 339
 Db 7 PheGluAsnAlaSerThrGlySerTyrThrAlaLeuIleSerLysGluGlyAspMetThr 26
 QY 340 -----AGTTTCTCTCGCGGAGATATCGATCGCGG 372
 Db 27 TyrGlyLeuAlaAspMetGluValPheAspTyrIleThrProGluPheLeuIleLysArg 46
 QY 373 GCGGAAGATTAAAAAATGCGAGCTTATTGTTCTGCAACTCGAAGTTTCAGCTTGAAACG 432
 Db 47 SerHisLeuLeuLysAlaLysCysIleIleValAspLeuAsnLeuGlyLysGluAla 66
 QY 433 GTTTATCACGCAATAGATTTCGCAAGAACACGCGGATTGAAGTGTATTAAAC----- 486
 Db 67 LeuAsnPheLeuCysAlaTyrThrThrLysHisGlnIleLysLeuValIleThrThrVal 86
 QY 487 CTGCGCGCAGCATTCGCGGAATTA---GATATGCTCTTATGCTGTAAATGCGATTCTTT 543
 Db 87 SerSerProLysMetLysAsnMetProAspSerLeuHisAla-----IleAspTrpIle 104

QY 544 GTACCTAATGAACCGAGCTGGAATATTAACCGGTATGCCAGTGGATACCTATGACCAT 603
 Db 105 IleThrAsnLysAspGluThrGluThrTyrLeuAsnLeuLysIleGluSerThrAspAsp 124
 QY 604 ATTCCGCGAGCGCGCGATTCCTGCTAGATAAAGGCTGGAACAATATTATTGTTCACCATG 663
 Db 125 LeuLysIleAlaAlaLysArgTipAsnAspLeuGlyValLysAsnValIleValThrAsn 144
 QY 664 GGCAGAAAGCGCGCTGTGGATACCGGTGACCAAGGAGTCCATGTTCCGCGCTTTAGA 723
 Db 145 GlyValLysGluLeuIleTyrArgSerGlyGluGluGluIleIleLysSerValMetPro 164
 QY 724 GTGACCGCTGT---GATACCGCGCGCGCGCATGCTTTATCGGCTGTTTCGCGCAT 780
 Db 165 SerAsnSerValLysAspValThrGlyAlaGlyAspPheCysAlaAlaValValTyr 184
 QY 781 TACTACGTCCAGAGCGGGATGTGGAAGCCGCCATGAAAAAAGCGCTCTCTTTGCCGCT 840
 Db 185 SerTrpLeuAsnGlyMetSerThrGluAspIleLeuIleAlaGlyMetValAsnAlaLys 204
 QY 841 TTCAGCGTCACCGGAAAGGACCC-----CAATCCTCTTTATCCA 879
 Db 205 LysThrIleGluThrLysTyrThrValArgGlnAsnLeuAspGlnGlnLeuTyrHis 224
 QY 880 AGCATTGAGCAATT 894
 Db 225 AspMetGluAspTyr 229
 RESULT 4
 US-10-485-517-185
 ; Sequence 185, Application US/10485517
 ; Publication No. US20050256299A1
 ; GENERAL INFORMATION:
 ; APPLICANT: University of Sheffield
 ; APPLICANT: Biosynexus Incorporated
 ; APPLICANT: Foster, Simon
 ; APPLICANT: Mond, James
 ; TITLE OF INVENTION: Antigenic Polypeptides
 ; FILE REFERENCE: P100629W0
 ; CURRENT APPLICATION NUMBER: US/10/485,517
 ; PRIOR FILING DATE: 2004-02-02
 ; PRIOR APPLICATION NUMBER: GB 0118825.9
 ; PRIOR FILING DATE: 2001-08-02
 ; PRIOR APPLICATION NUMBER: GB 0200349.9
 ; PRIOR FILING DATE: 2002-01-09
 ; NUMBER OF SEQ ID NOS: 424
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 185
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 US-10-485-517-185
 Alignment Scores:
 Pred. No.: 4,46e-06 Length: 306
 Score: 141.50 Matches: 65
 Percent Similarity: 39.19% Conservative: 51
 Best Local Similarity: 21.96% Mismatches: 117
 Query Match: 8.90% Indels: 63
 DB: 6 Gaps: 12
 US-10-049-750-11 (1-921) x US-10-485-517-185 (1-306)
 QY 31 GTGGACCTTATCACCTACCAACAGATGCCCAAGAGGGGAACTCTGGAAGCGCG 90
 Db 11 IleAspTyrValIlePheThrAsnAspPheLysIleAspGlyLeuAsnArgAlaThrAla 30
 QY 91 CGGTTTAAATCGGCTGCGCGGAGAAAGCGGCAACCGAGCGCTGGCGCGCGCTAAGCTC 150
 Db 31 ThrTyrLysPhe---AlaGlyGlyLysGlyIleAsnValSerArgValLeuLysThrLeu 49
 QY 151 AATTCAAAGTATTGATGTTGACCAAGTGGCG-----GACGATATTTTTCGCGACAAC 204


```
Db 1679 SerSerProLeuLysSerValValSerProValLysSerArgValAspValIleSerSer 1698
QY 306 -----
Db 1699 AlaLysIleThrMetAlaSerSerLeuSerSerProValLysGlnMetProGlyHisAla 1718
QY 305 GAGTTGGCGTTGACGAAATCGGCGCTACGCGCGTG-----CTG 267
|||:::|||||
Db 1719 GluValAlaLeuValAsnGlySerIleSerProLeuLysTyAlaSerSerSerThrLeu 1738
|||:::|||||
QY 266 GTACACGGTACTTTTCTACATACGCTGATTTGATCCCGAGATTCAGATTAGCAATG 207
|||:::|||||
Db 1739 IleAsnGlyCysLysAlaThr-----Ala 1746
|||:::|||||
QY 206 GTGTTGTGCGCAAAATATGTCGCCACT-----TTGGTCAACATCAAT 162
|||:::|||||
Db 1747 ThrLeuGlnGluLysIleSerSerAlaThrAsnSerValSerSerValIleSerAlaAla 1766
|||:::|||||
QY 161 ACTTTTGAATTCAGCTTAGCGCGCCGACGCGCTGTTTCGCCCTTTTCCGCCG----- 108
|||:::|||||
Db 1767 ThrAspThrValGluLysValPheSerThrThrThrAlaMetProPheSerProLeuArg 1786
|||:::|||||
QY 107 -----CAGCGGATTTTAAACGCCGCGCGCTTCCAGA 78
|||:::|||||
Db 1787 SerTyrValSerAlaAlaProSerAlaPheGlnSerLeuArgThrProSerAlaSerAla 1806
|||:::|||||
QY 77 GTTTCCTTCTTTGGGC 60
|||:::|||||
Db 1807 LeuTyrThrSerLeuGly 1812
|||:::|||||
RESULT 8
US-10-793-626-448
; Sequence 448, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 448
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-448
Alignment Scores:
Pred. No.: 0.232 Length: 187
Score: 91.00 Matches: 33
Percent Similarity: 43.66% Conservative: 29
Best Local Similarity: 23.24% Mismatches: 56
Query Match: 5.72% Indels: 24
DB: 6 Gaps: 5
US-10-049-750-11 (1-921) x US-10-793-626-448 (1-187)
QY 415 GAAGTTACGTTGAACCGGTTTATCAGCAATAGATTGGCAAG---AAACACGGGATT 471
|||:::|||||
Db 32 GluValLeuLeuGlu-----GluPheSerArgLeuAsnGluLysThr 45
|||:::|||||
QY 472 GAAGTTATTAAACCTTCGCCGACATTACGGGAATTAGATATGCTTATGCTCTGTAATA 531
|||:::|||||
Db 46 ThrIleIlePheAspProSerProArgIleAsnLysMetAsn----- 59
|||:::|||||
QY 532 TGCATTTCTTT-----GTACCTAATGAAACCGAGCTGGAA 567
|||:::|||||
```

```
Db 60 CysGluSerIleArgLysLeuLeuGluIleAsnThrIleValHisAlaAsnGluGlyGlu 79
QY 568 ATATTAAACCGGTATGTCAGTGTATACCTATGACCATATTCCGCGACGCGCACGTTCCGCTG 627
|||:::|||||
Db 80 IleLeu-----GlnLeuSerSerGluAsnHisValLysAspAlaIleLeuGluVal 96
|||:::|||||
QY 628 GTAGATAAAGGCGCTGAACAATATTATTGTCACCATGGCGGAGAAAGCGCCGCTGTGGATG 687
|||:::|||||
Db 97 SerLysGlnThrAsnGlnProValIleValThrLeuGlyAsnLysGlyThrLeuIleAla 116
|||:::|||||
QY 688 ACGGTGACACAGGAAGTCCATGTTCCGCGCTTTAGAGTGAACGCTGTGTGATACACGCGGC 747
|||:::|||||
Db 117 AsnLysCysLysValLysIleLeuGluGlyGluLysValProValThrAspThrIleGly 136
|||:::|||||
QY 748 GCGGCGCATGCTTTTATCGGCTGTTTCGCGCATTTACTACGTCCAGAGCGGGATGTGAA 807
|||:::|||||
Db 137 AlaGlyAspSerHisThrAlaAlaPheIleAlaGlyLeuLeuAspAsnGlnSerIleGlu 156
|||:::|||||
QY 808 GCCGCC 813
|||:::|||||
Db 157 LysAla 158
|||:::|||||
RESULT 9
US-10-467-657-3254
; Sequence 3254, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 3254
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-3254
Alignment Scores:
Pred. No.: 0.418 Length: 320
Score: 89.00 Matches: 58
Percent Similarity: 35.23% Conservative: 41
Best Local Similarity: 20.64% Mismatches: 140
Query Match: 5.60% Indels: 42
DB: 6 Gaps: 9
US-10-049-750-11 (1-921) x US-10-467-657-3254 (1-320)
QY 7 ATCCGCGTTATGGCTCTTAACATGTTGACCTTATCACCTACACCAACAGATGCCAAA 66
|||:::|||||
Db 19 ValLeuValValGlyAspValMetLeuAspArgTyrTrpPhe---GlyAspValSerArg 37
|||:::|||||
QY 67 GAAGGGGAAACTCTGGAAGCGCGCGGTTTAAATCGGC-----TGCGGCGGA 114
|||:::|||||
Db 38 IleSerProGluAlaProValProValAlaLysIleGlyArgIleAspGlnArgAlaGly 57
|||:::|||||
QY 115 AAAGGGCGAACACAGGCGTGGCGCGCTGAAGCTCAATTCAAAAGTTATGTTGACC 174
|||:::|||||
Db 58 GlyAlaAlaAsnValAlaArgAsnIleAlaSerLeuGlyArgAlaGlyLeuLeuSer 77
|||:::|||||
QY 175 AAAGTGGCGGACGAT-----ATTTTTCGCGACCAACACCAT 210
|||:::|||||
Db 78 ValThrGlyAspGluAlaAlaAlaLeuAspAlaLeuMetValGlnAspGlyVal 97
|||:::|||||
QY 211 CGTAAT-----CTCGAATCTCGGGGATCAATACGAGTATGTAGAAAAAGTACCGTGT 264
|||:::|||||
```

```

Db 98 AlaserTyrLeuMetArgAspLysGlnIleAlaThrThrValLysLeuArgValAla 117
QY 265 ACCACGAGCGGGTACGCGCGATTTCTCAACGCGCAACTCCAGCAACAGCATCTCGATC 324
Db 118 ArgAsnGlnIleLeuArgLeuAspPheGluGluHisProAsnArgLysValLeu--- 136
QY 325 ATCAAAGCGCTAACAGATTTCTTCGCGGGAAGATATCGATCGCGCG---CGCGAAGAT 381
Db 137 -----GluGlnIleLysArgLysArgGluIle 146
QY 382 TTAATAAATGACCATTTATTTCTGCAA-----CTGGAGTTTCAGCTTCAACGGTT 435
Db 147 LeuProGluTyrAspAlaIleIlePheSerAspTyrGlyLysGlyLeuSerHisIle 166
QY 436 TATCAGCAATAGAAATTTGGCAAGAAACACAGCGGATTGAAGTGTATTAAACCTCGCGCA 495
Db 167 SerAspMetIleAspTrpAlaLysHisValGlyLysThrValLeuIleAspPro----- 184
QY 496 GCATTTACGGGAATTAGATATGCTTTATGCTGCTGTAATGCGATTCTTTGTACCTAATGAA 555
Db 185 -----LysGlyAspAspTyrGluLysTyrValGlyAlaThrLeuIleThrProAsnCys 202
QY 556 ACCGAGCTGGAATATTAAACCGGTATGCCAGTGGATACCTATGACCATATTTCGCCGACGG 615
Db 203 AlaGluLeuLysGluValValGlySerTrpLysAsnGluGlyAspLeuThrGluLysAla 222
QY 616 GCACGCTGCTGTAGATAAAGGCTGAACATATTATTGTCCACCATGGCGGCAAGAGGC 675
Db 223 GlnAsnLeuArgArgHisLeuAspLeuThrAlaValLeuLeuThrArgSerGluGluGly 242
QY 676 GCGTGTGTGATGACGCGTACACGAGAGTCCATGTTCCGCGGTTTGTAGTGAACGCTGT 735
Db 243 MetThrLeuPheSerGluGlyGluProIleTyrGlnProThrArgAlaGlnGluValTyr 262
QY 736 GATACAGCGGCGCGCGGATGCTTTATC-----GGC 768
Db 263 AspValSerGlyAlaGlyAspThrValIleAlaGlyMetGlyLeuGlyLeuAlaLagly 282
QY 769 TGT 771
Db 283 Cys 283

```

RESULT 10

US-10-858-730-209

; Sequence 209, Application US/10858730

; Publication No. US2005025568A1

; GENERAL INFORMATION:

; APPLICANT: Bailey, Richard B.

; APPLICANT: Blomquist, Paul

; APPLICANT: Doten, Reed

; APPLICANT: Driggers, Edward M.

; APPLICANT: Madden, Kevin T.

; APPLICANT: O'Leary, Jessica

; APPLICANT: O'Toole, George

; APPLICANT: Trueheart, Joshua

; APPLICANT: Walbridge, Michael J.

; APPLICANT: Yorgsey, Peter S.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID

; FILE OF INVENTION: PRODUCTION

; FILE REFERENCE: 14184-030001

; CURRENT APPLICATION NUMBER: US/10/858,730

; CURRENT FILING DATE: 2004-06-01

; PRIOR APPLICATION NUMBER: US 60/475,000

; PRIOR FILING DATE: 2003-05-30

; PRIOR APPLICATION NUMBER: US 60/551,860

; PRIOR FILING DATE: 2004-03-10

; NUMBER OF SEQ ID NOS: 364

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 209

; LENGTH: 445

; TYPE: PRT

; ORGANISM: Coryne-bacterium glutamicum

US-10-858-730-209

Alignment Scores:

```

Pred. No.: 0.636 Length: 445
Score: 87.50 Matches: 64
Percent Similarity: 38.32% Conservative: 41
Best Local Similarity: 23.36% Mismatches: 88
Query Match: 5.50% Indels: 81
DB: 6 Gaps: 16

```

US-10-049-750-11 (1-921) x US-10-858-730-209 (1-445)

```

QY 85 CGCGCGCGTTTAAATCGGTCGCGGGAAGAGGCGAACCAGCGCGTGTGCGGCGCT 144
Db 6 AlaProSerPheAsn-----ProGlyLysGlyProGlySerAlaValGlyLeuAla 22
QY 145 AAGCTC-----AATTCAAAGTATTG-----ATGTTGACCAAAAGTGGCGAC 186
Db 23 LeuLeuGlyPheGlyThrValGlyThrGluValMetArgLeuMetThrGluTyrGlyAsp 42
QY 187 GATATTTTTCGCGCAACACCATTCGTAATCTCGAATCCTGGGATCCTACGACGTAT 246
Db 43 GluLeu---AlaHisArgIleGlyGlyProLeuGluValArgGlyIleAlaValSerAsp 61
QY 247 GTAGAAAAGTACCGTGTACGAGCAGCGCGTACGCGCGATTTCGTCAACGCCCACTCC 306
Db 62 IleSerLys-----ProArgGluGlyValAlaPro----- 71
QY 307 AGCAACAGCATTCGATCATCAAAGCGCTAACAGTTTCTCTCCCGGAAGATATCGAT 366
Db 72 -----GluLeuLeuThrGluAspAlaPheAlaLeuIleGluArgGluAspValAsp 88
QY 367 CGCGCGCGGAAGATTAAATAAATGCGCAGCTTATTGTTCTGCAACTCGAAGTTCAGCTT 426
Db 89 IleValValGlu-----ValIleGlyGlyIleGluTyrProArg 101
QY 427 GAAACGTTTATCACGCAATAGAAATTTGGCAAGAA----- 462
Db 102 GluValValLeuAlaLeuLeuLysAlaGlyLysSerValValThrAlaAsnLysAlaLeu 121
QY 463 -----CACGGGATTGAAGTGTATTATAAACCTCGCGCCAGCATTACGGGAATTAGAT 513
Db 122 ValAlaAlaHisSerAlaGluLeu-----AlaAsp 131
QY 514 ATGTTCTTATGCTGTAAATGCGATTCTTTGTACCTAATGAACCGAGCTGGAATATTA 573
Db 132 AlaAlaGluAlaAlaAsnValAspLeuTyrPhe-----GluAlaAlaValAla 147
QY 574 ACCGTATGCGATGCGATACCTATGACCATATTCCGCGAGCGGCACGTTCCGTTGTAGAT 633
Db 148 GlyAlaIleProVal-----ValGlyProLeuArgArgSerLeuAlaGly 162
QY 634 AAAGGCTGAACAATATTATTGTCCATGCGCGAGAAAGCGCGCTGTGATGACGCGT 693
Db 163 AspGlnIleGlnSerValMet-----GlyIleValAsnGlyThrThr 176
QY 694 GACCAGGAAGTCCATGTTCCGCGGTTTAGAGTGAACGCTGTTGTATACACGCGCGCG 750
Db 177 Asn-----PheIleLeuAspAlaMetAspSerThrGlyAlaAsp 189
QY 751 ---GCGGATGCTTTTATCGGCTGTTTCGGGCATTACTACGTCAG-----ACGCGG 798
Db 190 TyrAlaAspSerLeuAlaGluAlaThrArgLeuGlyThrAlaGluAlaAspProThrAla 209
QY 799 GATGTGGAA-----GCCGCCATGAAAGCGCGCTCCTC 831
Db 210 AspValGluGlyHisAspAlaAlaSerLysAlaAlaIleLeu 223

```

RESULT 11

US-10-467-7660

; Sequence 7660, Application US/10467657

; Publication No. US20050260581A1

; GENERAL INFORMATION:

```
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 7660
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-7660

Alignment Scores:
Pred. No.: 0.901 Length: 482
Score: 86.00 Matches: 75
Percent Similarity: 33.15% Conservative: 43
Best Local Similarity: 21.07% Mismatches: 118
Query Match: 5.41% Indels: 120
DB: 6 Gaps: 17

US-10-049-750-11 (1-921) x US-10-467-657-7660 (1-482)

Qy 7 ATCCGGTATTGGCTTAACATGGT-----GACCTTATC 42
Db :|||||:|||||:|||||:|||||:
10 LeuAlaValMetGlyGlnAsnLeuLeuAsnMetAsnAspCysGlyPheLysValVal 29
Qy 43 ACCTACACCAACAGATCCCAAGAGAGGAACTCTGGAGGCGCGCGCTTTAAATC 102
Db :|||:|||||:|||||:|||||:
30 AlaTyrAsnAArgThrIleGlyLysValAspGluPheLeuAsnGlyAlaAlaLysGluThr 49
Qy 103 GGC---TCGCGGGNAAGGGCGCAACAGCCGCGCGCTTAAGCTCAATTCACAAA 159
Db :|||||:|||||:|||||:|||||:
50 GlyIleIleGlyAlaTyrSerLeuGlnAspLeuValAspLysLeuAlaLysProArgLys 69
Qy 160 GTATTGATGTTGACCAAGTGGCGAC-----GATATTTTTCGCCGACACACCATTCGT 213
Db :|||||:|||||:|||||:|||||:
70 IleMetMetValArgAlaGlySerValValAspPheValGluGlnLeuPro 89
Qy 214 AATCTCGAATCCTGGGGATC-----AATACGACGTATGTAGAAAAAGTA 258
Db :|||||:|||||:|||||:|||||:
90 LeuLeuGluGluGlyAspIleIleAspGlyAsnAlaAsnTyrProAspThrThr 109
Qy 259 CCGTGTACCAGC-----AGCGCGTAGCGCGATTTTCGTCACAGCCCAACTCCAGCAAC 312
Db :|||||:|||||:|||||:|||||:
110 ArgArgThrHisTyrLeuAlaGlyLysGlyIleLeuPheValGlyAlaGlyValSerGly 129
Qy 313 AGCATTCGATCATCAAGGGCGCTAACAG-----TTTCTCTCGCCCGGAAGATATCGAT 366
Db :|||||:|||||:|||||:|||||:
130 Gly-----GluGluGlyAlaArgGlyProSerIleMetProGlyGlyAspLys 146
Qy 367 CGCGCGCGGAGATTTAAAAAATGCGCAGCTTATGTTCTGCAACTCGAAGTTTCAG--- 423
Db :|||||:|||||:|||||:|||||:
147 ArgAlaTyrGluAlaValLysProIlePheGlnAlaIleAlaAlaLysThrProGlnGly 166
Qy 424 -----CTTGAACCGTTTAT 438
Db :|||||:|||||:|||||:|||||:
167 GluProCysCysAspTrpValGlyLysAspGlyAlaGlyHisPheValLysMetValHis 186
Qy 439 CACCAATAGAAATTTGGCAAGAAACACGGGATTTGAAGTGTATTAAACCCCTGCGCCAGCA 498
Db :|||:|||||:|||||:|||||:
187 AsnGlyIleGluTyrGly----- 192
Qy 499 TTACGGGAATTAGATATCTTATGCCCTGTAATTCGGATTTCTTTGTACCT----- 549
Db :|||||:|||||:|||||:|||||:
193 -----AspMetGlnLeuIleCysGluAlaTyrGlnPheMetLysAspGlyLeu 208
```

```
Qy 550 -----AATGAAACCGAGCTGAA 567
Db :|||||:|||||:|||||:|||||:
209 GlyLeuSerTyrAspGluMetHisArgValPheAlaGluTrpAsnLysThrGluLeu--- 227
Qy 568 ATATTAAACCGTATGCCAGTGGATACCTAT---GACCATATTCCGCGAGCGGACGTTTCG 624
Db :|||||:|||||:|||||:|||||:
228 -----AspSerTyrLeuIleGluIleThrAlaAlaIleLeuGly 240
Qy 625 CTGCTAGATAAAGG-----CTGAACAATATTATTGTCAACATGGCGGAGAAA 672
Db :|||||:|||||:|||||:|||||:
241 TyrLysAspGluGlyGlyGluProLeuAlaGluLysIleLeuAspThrAlaGlyGlnLys 260
Qy 673 GCGCGCGTGTGGATGAGCGCGTACCAGGAAGTCCATGTTCCGGCGTTTAGAGTGAACGCT 732
Db :|||||:|||||:|||||:|||||:
261 GlyThrGlyLysTrpThr-----GlyIleAsnAla 270
Qy 733 GTTGAT-----ACCAGCGCGCGCGGATGCTTTATCGGCTGT----- 771
Db :|||||:|||||:|||||:|||||:
271 LeuAspLeuGlyIleProLeuThrLeuIleSerGluAlaValPheAlaArgCysValSer 290
Qy 772 ---TTCGCGCATTAACGTCCAGCGCGGATGTGAAGCC----- 810
Db :|||||:|||||:|||||:|||||:
291 SerPheLysGluGlnArgValGlnThrGlyLysLeuPheAlaArgThrAlaThrProVal 310
Qy 811 -----GCCATGAAAAAGCCGCTCTC 831
Db :|||||:|||||:|||||:|||||:
311 GluGlyGlyLysGlnGluTrpValGluAlaLeuArgGlnAlaLeuLeu 326

RESULT 12
US-10-793-626-2530
; Sequence 2530, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2530
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2530

Alignment Scores:
Pred. No.: 0.856 Length: 186
Score: 85.00 Matches: 28
Percent Similarity: 41.32% Conservative: 22
Best Local Similarity: 23.14% Mismatches: 59
Query Match: 5.35% Indels: 12
DB: 6 Gaps: 3

US-10-049-750-11 (1-921) x US-10-793-626-2530 (1-186)

Qy 535 GATTTCCTTTGTACCTAATGAACCGAGCTGGAAATATTAAACGGTATGCCAGTGGATACC 594
Db :|||||:|||||:|||||:|||||:
48 AspValValThrProAsnIleProGluAlaGluGluIleThrGlyIleLysIleAsnAsp 67
Qy 595 TATGACCATATTCGCGCAGCGGACGTTCCGCTGGTAGATAAA---GGCTGAACATATT 651
Db :|||:|||||:|||||:|||||:
68 GluGluSerIleArgLysAlaGlyGlnIlePheIleAsnGluIleGlySerLysGlyVal 87
Qy 652 ATTGTCACCATGGCGAGAAAGCGCGCTG-----TGGATGACGCGT 693
Db :|||||:|||||:|||||:|||||:
88 ValIleLysGlyGlyHisSerAlaAspLeuAsnAlaLysAspPheLeuPheThrLys 107
```


Qy 694 GACCAGGAAGTCCATGTTCCGCGCTTTAGAGTGAACGCTGTGTGATACACGCGCGCGGC 753
Db 108 AsnGluThrTyrThrPheGluAsnLysArgPheAspThrLysHisThrHisGlyThrGly 127
Qy 754 GATGCTTTATCGGCTGTTCCGCGATTACTACGTCACAGCGGGATGTGGAAGCCGCC 813
Db 128 CysThrPheSerAlaValIleThrAlaGluLeuAlaLysGlyArgSerIleLysAspAla 147
Qy 814 ATGAAAAAGCGCTCTCTTTCGCCCTTTCAGCGTC-----ACCGGGAAA 858
Db 148 ValLysLysAlaLysGluPheIleSerLeuSerIleGluHisThrProGluIleGlyLys 167
Qy 859 GGC 861
Db 168 Gly 168

RESULT 13

US-11-055-822-170
; Sequence 170, Application US/11055822
; Publication No. US20050260707A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
; TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS
; FILE REFERENCE: BGI-12ICPCN
; CURRENT APPLICATION NUMBER: US/11/055,822
; CURRENT FILING DATE: 2005-02-11
; PRIOR APPLICATION NUMBER: 09/606,740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141,031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/142,101
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: 60/148,613
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/187,970
; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: DE 19930476.9
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: DE 19931415.2
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931418.7
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931419.5
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931420.9
; PRIOR FILING DATE: 1999-07-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; SEQ ID NO 170
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-11-055-822-170

Alignment Scores:
Pred. No.: 1.52 Length: 445
Score: 83.50 Matches: 61
Percent Similarity: 37.59% Conservatives: 42
Best Local Similarity: 22.26% Mismatches: 90
Query Match: 5.25% Indels: 81
DB: 7 Gaps: 15

US-10-049-750-11 (1-921) x US-11-055-822-170 (1-445)

Qy 85 GCGCGCGCTTTAAATCGGCTGCGCGGAAAGGCGCAACGAGCGCTGCGCGCGCT 144
Db 6 AlaProSerPheAsn-----ProGlyLysGlyProGlySerAlaValGlyIleAla 22

Qy 145 AAGCTC-----AATTCAAAAGTATTG---ATGTTGACCAAAAGTGGCGAC 186
Db 23 LeuLeuGlyPheGlyThrValGlyThrGluValMetArgLeuMetThrGluThrGlyAsp 42
Qy 187 GATATTTTGGCGCAACACCATTCGTAATCTCAATCTCGGGGATCAATACGACGAT 246
Db 43 GluLeu---AlaHisArgIleGlyGlyProLeuGluValArgGlyIleAlaValSerAsp 61
Qy 247 GTAGAAAAAGPACCGTGTATCCAGCAGCGCGCTAGCGCGATTTTCGTCACGCCCACTCC 306
Db 62 IleSerLys-----ProArgGluGlyValAlaPro----- 71
Qy 307 AGCAACGATTCGATCATCATCAAGGCGCTAAAGTTTCTCTCGCCGGAAGATATCGAT 366
Db 72 -----GluLeuLeuThrGluAspAlaPheAlaLeuIleGluArgGluAspValAsp 88
Qy 367 CGCGCGGGAAGATTTAAAAAATGCCAGCTTATTGTTCTGCAACTGGAAAGTTCAGCTT 426
Db 89 IleValValGlu-----ValIleGlyGlyIleGluThrProArg 101
Qy 427 GAAACGGTTTATCACGCAATAGAAATTTGGCAAGAAA----- 462
Db 102 GluValValLeuAlaAlaLeuLysAlaGlyLysSerValValThrAlaAsnLysAlaLeu 121
Qy 463 -----CACGGGATTGAAGTGTATTAAACCTCGCCGACGATTACGGGAATTAGAT 513
Db 122 ValAlaAlaHisSerAlaGluLeu-----AlaAsp 131
Qy 514 ATGCTTATGCTGTAAATGCGATTTCCTTGTACCTAATGAACCGAGCTGGAATATTA 573
Db 132 AlaAlaGluAlaAlaAsnValAspLeuThrPhe-----GluAlaAlaValAla 147
Qy 574 ACCGGTATGCGCAGTGGATACCTATGACCATATTCGCGGAGCGGACGTTGCTGTAGAT 633
Db 148 CysAlaIleProVal-----ValGlyProLeuArgArgSerLeuAlaGly 162
Qy 634 AAAGGCTGAACATATTTGTCACCATGGCGGAAGCGCGCTGTGGATGACGCT 693
Db 163 AspGlnIleGlnSerValMet-----GlyIleValAsnGlyThrThr 176
Qy 694 GACCAGGAAGTCCATGTTCCGCGCTTTAGAGTGAACGCTGTGTATACACGCGCGCG--- 750
Db 177 Asn-----PheIleLeuAspAlaMetAspSerThrGlyAlaAsp 189
Qy 751 ---GCGGATGCTTTATCGGCTGTTTCGCGCATCTACTACGTCACAGCGGGGATGTG--- 804
Db 190 TyrAlaAspSerLeuAlaGluAlaThrArgLeuGlyTyrAlaGluAlaAspProThrAla 209
Qy 805 -----GAAGCGCCCATGAAAAAGCGCTCCTC 831
Db 210 AsnValGluGlyHisAspAlaAlaSerLysAlaAlaIleLeu 223

RESULT 14
US-10-467-657-3330
; Sequence 330, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 3330
; LENGTH: 244
; TYPE: PRT

```
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-3330

Alignment Scores:
Pred. No.: 1.43 Length: 244
Score: 83.00 Matches: 51
Percent Similarity: 35.47% Conservative: 21
Best Local Similarity: 25.12% Mismatches: 68
Query Match: 5.22% Indels: 63
DB: 6 Gaps: 10

US-10-049-750-11 (1-921) x US-10-467-657-3330 (1-244)
QY 363 CGATCGCGCGCGAAGATTTAAA-----AAAATCCAGCTTAT 401
Db 82 LysSerLysAsnGlySerPheGlyThrGluLeuValSerArgSerAlaMetProArgTyr 101
QY 402 TGTTCGCAACTGGAAGTTTCAGCTTGAACCGTTTATCACCAATAGAAATTTGCAAGAA 461
Db 102 GlnTyrThrAsnGlyArgArgIleGlnThrGly-----TipGluGlu 115
QY 462 ACA-----CGGATTGAAGTGTATTAAACCCCTGCCGACGACATT 500
Db 116 ArgAlaGluPheLysAlaGluGlyArgAspPheAspAla-LeuAsnArgPheIleAlaAs 135
QY 501 ACGGGAATTAGATATGCTTATGCTGCTGTAATGCGATTCTTTGTACCTAATGAACCGA 560
Db 135 pValGlnThrAspAlaSerLeuGlu---AspThrAspPheSerValSerArgGluArgAr 154
QY 561 GCTGGAATATTAAACCGGTATGCCAGTGGATACCTATGACCATTATTCGCGCAGCGGCACG 620
Db 154 GAsnGluValIleAspGlnValSerLysAspAlaValLeuArgPheLysAlaArgAlaG1 174
QY 621 TTCGCTGTGTAGAT-----AAAGGCTGAACAAATATTATTGTCACCATGGGCGGA 668
Db 174 uLysLeuAlaGlyValLeuGlyAlaSerGlyTyrLysIleValLysLeuAsnPheGlyG1 194
QY 669 GAAAGGCGCGTGTGGATGACGCGTGACCGAAGTCCATGTCGGGGTTTAGAGTGAA 728
Db 194 nileGlySer-----HisIle----- 199
QY 729 CGCTGTTGATACCGCGCGCGCGATGCCCTTTATCGGCTGTTTTCGGCGCATTTACTAGT 788
Db 200 -----AlaGlyAsp----- 202
QY 789 CCAGAGCGGGATGTGGAAGCGCGCATGAAAGCCGCTCTCTTTGCGCGCTTTCAGCGT 848
Db 203 -----GlyAlaValArgAlaLysMetLeuArgAlaMetProMetAlaAla---SerVa 219
QY 849 CACCGGGAAGCCACCAATCCTCTTATCCAGCATTGAGCAATTTAATGAGTATCTTTC 908
Db 219 lAsnMetLysGlyThrAspSerAlaAlaProGlyValGluIleSer-----IleSe 237
QY 909 GTTGAAC 915
Db 237 xIleAsn 239
```

RESULT 15

```
US-10-467-657-2702
; Sequence 2702, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
```

```
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 2702
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-2702

Alignment Scores:
Pred. No.: 1.64 Length: 400
Score: 83.00 Matches: 40
Percent Similarity: 36.88% Conservative: 19
Best Local Similarity: 25.00% Mismatches: 53
Query Match: 5.22% Indels: 48
DB: 6 Gaps: 8

US-10-049-750-11 (1-921) x US-10-467-657-2702 (1-400)
QY 436 TATCACGCAATAGAA-----TTTGGCAAGAAACACGGGATTGAAAGTGTATTATA 483
Db 4 TyrHisAlaProAspGluLysGlyPhePheGlyGluHisGlyGlyLeuTyr-ValSerGlu 23
QY 484 AACCTCGCCGACGATTACGGGAATTAGATATGCTTAT---GCCTGTAATCGCAT--- 537
Db 24 ThrLeuIleProAlaLeuLysGluLeuGlnAlaTyrAsnGluAlaLysAsnAspPro 43
QY 538 ---TTCCTTGTACCTAATGAAACCGAGCTGAAATATTAAACCGGTATGCCAGTGGATACC 594
Db 44 GluPheTrpAlaGluPheArgArgAspLeuLysHisTyrValGlyArgProSerProVal 63
QY 595 TATGACCATTATTCGCGCAGCGCACGTTTCGCTGGTAGATAAAGGCTGAACAATATTATT 654
Db 64 TyrHis-----AlaAlaArgLeuSerGluHisLeuGly----- 74
QY 655 GTCACCATGGCGGAGAAAGCGCG---CTGTGGATGACGCTGACCGAGGAAGTCCATGTT 711
Db 75 -----GlyAlaGlnIleTrpLeuLysArgGluAspLeuAsnHisThr 88
QY 712 CCGCGCTTTAGAGTGAAC----- 729
Db 89 GlyAlaHisLysValAsnAsnThrIleGlyGlnAlaLeuAlaArgArgMetGlyLys 108
QY 730 -----GCTGTGTATACCGCGCGCGCGCAT----- 756
Db 109 LysArgValIleAlaGluThrGlyAlaGlyGlnHisGlyValAlaSerAlaThrValAla 128
QY 757 GCCTTTATCGGCTGTTTCGCGCATTTACTAGCTCAGACGGGGATGTGGAAGCCGCGCATG 816
Db 129 AlaArgPheGlyMetThrCysAspValTyrMetGlyAlaAspAspIleGlnArgGlnMet 148

Search completed: December 23, 2005, 23:20:32
Job time : 16.5438 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 23, 2005, 22:20:36 ; Search time 40.015 Seconds
(without alignments)
4429.126 Million cell updates/sec

Title: US-10-049-750-11

Perfect score: 1590

Sequence: 1 atggatcgcgggtattgg.....atcttcgtgaacgaataa 921

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10049750/runat_23122005_113553_6303/app_query.fasta_1.1742
-DB=PIR -QFMT=FASTAN -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HRAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10049750 @CGN 1 1 83 @runat_23122005_113553_6303 -NCPU=3
-NO MMAP -LARGSEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR 80.*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1570	98.7	306	2	AE0963
2	878	55.2	312	2	AC3153
3	878	55.2	312	2	H98134
4	550	34.6	308	2	B83403
5	523	32.9	343	2	T20529
6	480.5	30.2	304	2	G89790
7	474.5	29.8	293	1	D69690
8	468.5	29.5	309	2	AB0952
9	468.5	29.5	318	2	T39594
10	463.5	29.2	308	2	AI0001
11	460	28.9	294	2	H84115
12	460	28.9	304	2	T13380
13	450.5	28.3	306	2	B64073
14	445.5	28.0	309	1	KIECRB

15	445.5	28.0	309	2	F91215
16	445.5	28.0	309	2	G86061
17	438	27.5	307	2	AB2029
18	434.5	27.3	306	2	D82497
19	434	27.3	404	2	AF0994
20	414	26.0	299	2	E72311
21	407.5	25.6	300	2	G86829
22	399.5	25.1	313	2	AG3520
23	396	24.9	293	2	C90139
24	382.5	24.1	310	2	AF0440
25	381	24.0	300	2	A75599
26	355.5	22.4	303	2	B87367
27	352.5	22.2	378	2	F86307
28	352	22.1	299	2	D97403
29	352	22.1	299	2	AD2621
30	297.5	18.7	333	1	K1BYRB
31	291.5	18.3	305	2	T44955
32	290.5	18.3	304	2	B70680
33	288.5	18.1	300	1	S40827
34	288.5	18.1	300	2	F91229
35	288.5	18.1	300	2	E86076
36	286.5	18.0	298	2	AH0947
37	247.5	15.6	331	2	T14544
38	245	15.4	317	2	B82815
39	244	15.3	298	2	AF2933
40	244	15.3	298	2	A98349
41	241.5	15.2	325	2	B84720
42	240	15.1	310	2	G72752
43	236.5	14.9	319	2	AC0942
44	233.5	14.7	319	2	A83882
45	233.5	14.7	345	2	C86195

ALIGNMENTS

RESULT 1

AE0963
probable carbohydrtase kinase STY3989 [imported] - Salmonella enterica subsp. enterica serov. typhi
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AB0963
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dord, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Party, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov. typhi
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AE0963
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-306 <PAR>
A:Cross-references: UNIPARC:UPI00000380F2; GB:AL513382; PIDN:CAD03201.1; PID:gi16504832;
C:Genetics:
A:Gene: STY3989
C:Superfamily: ribokinase

Alignment Scores:	Length:	Matches:
Pred. No.:	1.39e-129	306
Score:	1570.00	306
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	98.74%	Indels: 0
DB:	2	Gaps: 0

US-10-049-750-11 (1-921) x AE0963 (1-306)

QY	1	ATGGATATCGCGTTATTGGCTTAACATGGTGGACCTTATCACCTACCAACGAGATG 60
Db	1	MetAapIleAlaValIleGlySerAsnMetValAspLeuIleThrTyrThrAsnGlnMet 20
QY	61	CCCAAGAAGGGGAAACTCTGGAAGCGCGCGGCTTTAAATCGGCTGGCGCGGAAGGG 120

```
|||||
21  ProLysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyGlyLysGly 40
QY 121  GCGAACCAAGCGCGCGCGCTTAAGCTCAATTCAAAGATATTGATGTTGACCAAGTG 180
Db 41  AlaAsnGlnAlaValAlaAlaLysLeuAsnSerLysValLeuMetLeuThrLysVal 60
QY 181  GGCCACGATATTTTTCGCCGACAAACACCATTCCTGAATCTCGAATCTCGGGGATCAATACG 240
Db 61  GlyAspAspIlePheAlaAspAsnThrIleArgAsnLeuGluSerTrpGlyIleAsnThr 80
QY 241  ACGTATGTAGAAAAGTACCGTGTACCAAGCAGCGCGTAGCGCGATTTTCGTCAACGCC 300
Db 81  ThrTyrValGluLysValProCysThrSerSerGlyValAlaProIlePheValAsnAla 100
QY 301  AACTCCAGCAACAGCATTCGATCATCAAGCGCTTAACAAGTTCTCTCGCCCGAAGAT 360
Db 101  AsnSerSerAsnSerIleLeuIleLysGlyAlaAsnLysPheLeuSerProGluAsp 120
QY 361  ATCGATCGCGCGCGGAGATTTAAAAAATGCCAGCTTATTGTTCTGCACTCGAAGTT 420
Db 121  IleAspArgAlaAlaGluAspLeuLysCysGlnLeuIleValLeuGlnLeuGluVal 140
QY 421  CAGCTTGAACCGTTTATCACGCAATAGAAATTTGCGAAGAAACACGGGATTTGAAGTTA 480
Db 141  GlnLeuGluThrValTyrHisAlaIleGluPheGlyLysLysHisGlyIleGluValLeu 160
QY 481  TTAACCCCTCGCCAGCAATACCGGAATAGATATGCTTATGCTGTAATCGGATTC 540
Db 161  LeuAsnProAlaProAlaLeuArgGluLeuAspMetSerTyrAlaCysLysCysAspPhe 180
QY 541  TTTGTACTTAATGAACCGAGCTGGAATATTAACCGGTATGCCAGTGATGATACCTATGAC 600
Db 181  PheValProAsnGluThrGluLeuGluIleLeuThrGlyMetProValAspThrTyrAsp 200
QY 601  CATATTCCGCGAGCGGCACGTTTCGCTGCTAGATAAAGCGCTGAACAAATATTATTGTCCACC 660
Db 201  HisIleArgAlaAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnIleValThr 220
QY 661  ATGGCGGAGAAAGCGCGCTGTGATGACGCGTGACACGAGGAGTCCATGTTCCGCGCTT 720
Db 221  MetGlyGluLysGlyAlaLeuTrpMetThrArgAspGlnGluValHisValProAlaPhe 240
QY 721  AGAGTGAACGCTGTGATACCAAGCGCGCGCGCATGCTTTATCGGCTGTTTCGCGCAT 780
Db 241  ArgValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheAlaHis 260
QY 781  TACTACGTCCAGAGCGGGATGTGGAAGCCGCCATGAAAAAGCGCTCTCTTTGCGCGCT 840
Db 261  TyrTyrValGlnSerGlyAspValGluAlaAlaMetLysLysAlaValLeuPheAlaAla 280
QY 841  TTCAGCTCACCGGAAAGGACCAATCTCTTATCCAGCATGTAGCAATTTAATGAG 900
Db 281  PheSerValThrGlyLysGlyThrGlnSerSerTyrProSerIleGluGlnPheAsnGlu 300
QY 901  TATCTTTCGTGAACGAA 918
Db 301  TyrLeuSerLeuAsnGlu 306

RESULT 2
AC3153
ribokinase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
: Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
```

```
A:Accession: AC3153
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-312 <KUR>
A:Cross-references: UNIPROT:Q8UG60, UNIPARC:UPI000000D2715, GB:AE008689; PIDN:RAL45641.1;
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: rbsK
A:Map position: linear chromosome
C:Superfamily: ribokinase

Alignment Scores:
Pred. No.: 7,41e-69 Length: 312
Score: 878.00 Matches: 168
Percent Similarity: 76.25% Conservative: 60
Best Local Similarity: 56.19% Mismatches: 71
Query Match: 55.22% Indels: 0
DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x AC3153 (1-312)
```

```
QY 7 ATCGCGTTATTGGCTCTAACATGCTGACCTTATCACCTACACCAACAGATGCCCAA 66
Db 5 IleGlyValValGlySerAsnMetValAspLeuIleThrTyrValAspArgMetProGly 24
QY 67 GAAGGGGAAAATCTCTGGAAGCGCGCGCTTTAAATTCGGCTCGCGCGGAAAAGGGCGAAC 126
Db 25 ProGlyGluThrLeuGluAlaProThrPheGluMetGlyCysGlyGlyLysGlyAlaAsn 44
QY 127 CAGCGCTGCGCGCGCTTAAGCTCAATCAAAAGTATTGATGTTGACCAAGTGGCGAC 186
Db 45 GlnAlaValAlaAlaAlaArgLeuGlyAlaGluValMetMetValThrArgValGlyAsp 64
QY 187 GATATTTTCCGCAACACCATTCGTAATCTCGAATCTCGGGGATCAATPACACGTAT 246
Db 65 AspValPheAlaAspAsnThrIleArgAsnLeuAlaSerPheGlyValAspThrArgHis 84
QY 247 GTAGAAAAGTATCCGTGTACCAAGCGCGGTAGCGCGATTTTCGTCAACGCCCACTCC 306
Db 85 ValValLysValSerGlyLysSerSerGlyValAlaProIlePheValGluGlnSerGly 104
QY 307 AGCAACAGCATTCGATCATCAAGGCGCTTAACAGTTTCTCTCGCGGAGATATCGAT 366
Db 105 GluAsnSerIleLeuIleValLysGlyAlaAsnAlaAspLeuLeuProValGluValAsp 124
QY 367 CGCGCGCGGAGATTTAAATAATGCCAGCTTATTGTTCTGCAACTGGAAGTTTCAGCTT 426
Db 125 LysAlaAlaAlaAspLeuLysGluCysGlyLeuIleLeuMetGlnMetGluValProVal 144
QY 427 GAAACGCTTTATCACGCAATAGAAATTTGGCAAGAAACACACGGGATTTGAAGTGTATTAAAC 486
Db 145 GluThrValTyrHisThrIleGluPheAlaAlaGlnAsnGlyIleGluThrIleLeuAsn 164
QY 487 CTGCGCGCAGCATTTACCGGGAATTAGATGTCTTATGCTGCTGTAATGCGATTTCTTTGTA 546
Db 165 ProAlaProAlaAlaAlaAsnLeuAspProGluArgIleArgGlnValThrPheLeuVal 184
QY 547 CCTAATCAACCGAGCTGGAAATATTAAACCGTATGCCAGTGATACCTATGACCATATT 606
Db 185 ProAsnGluSerGluLeuAlaLeuLeuSerGlyLeuProThrAspThrAspGluAspIle 204
QY 607 CGCGCAGCGCACGCTTCGCTGCTAGATAAAGGCTGAACATATTATTGTCCATCGGC 666
Db 205 ValArgAlaAlaArgSerLeuIleAlaArgGlyIleArgThrValIleValThrLeuGly 224
QY 667 GAGAAAGCGCGCTGTGTGATGACCGCTGACACGAAAGTCCATGTTCCGGCGTTTAGAGTG 726
Db 225 AlaArgGlyAlaArgMetIleThrSerAspGluIleValAsnIleGluProValLysVal 244
QY 727 AACGCTGTTGATACACGCGCGCGCGATGCCCTTTATCGGCTGTTTCGCCCATTTACVAC 786
Db 245 ThrProArgAspThrThrGlyAlaGlyAspAlaPheIleGlySerPheAlaArgPheTyr 264
```



```
Qy 247 GTAGAAAAAGTACCGTGTACCAGCGCGGTAGCGCGATTTTCGTCAACGCCAACTCC 306
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85 ValGluArgValAlaGlyGluSerSerGlyValAlaLeuIleValValAspSerSer 104
Qy 307 AGCAACAGCATTCGTGATCATCAAAAGCGCTAACAAAGTTTCTCTCGCGGAAGATATCGAT 366
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 105 GlnAenAlaIleValIleValAlaGlyGlyAenGlyHisLeuSerProAlaValLeuAla 124
Qy 367 CGCGCGGCGGAAGATTAAAAAATGCCAGCTATTGTTCTGCACTGGAAGTTCAGCTT 426
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 125 ArgHisGluHisLeuLeuGluGlnAlaGlnValValValCysGlnLeuGluSerProLeu 144
Qy 427 GAAACGGTTTATCAGCAATAGATTTCGGCAAGAAACACACGGGATTGGAAGTCTTATTAAAC 486
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 145 GluThrValGlyHisValLeuArgArgAlaHisAlaLeuGlyLysThrValIleLeuAen 164
Qy 487 CTCTGCGCAGCATTAACGGGAATTAGATATGCTTATGCTGCTGTAATGCGATTTCTTTGTA 546
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 165 ProAlaProAlaThrArgAspValProAlaGluTrpLeuProLeuValAspTyrLeuVal 184
Qy 547 CCTAATGAACCGAGCTGGAATATTAAACGGTATGCCAGTATGCCAGTATGACCATATT 606
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 185 ProAenGluThrGluSerGluLeuLeuCysArgLeuProValAspSerLeuGluSerAla 204
Qy 607 CGCGCAGCGCAGCTGCTGCTAGATAAAGCGCTGAACAATATTATTGTCACCATGGC 666
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 205 GlyArgAlaAlaGluArgLeuGluGluMetGlyAlaGlyArgValIleValThrLeuGly 224
Qy 667 GAGAAAGCGCGCTGTGGATGACGCGTACACGGAAGTCCATGTTCCGCGCTTTAGAGTG 726
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 225 AlaGlnGlyAlaLeuLeuValGlyGluGlyArgValGluHisPheProValAlaArgVal 244
Qy 727 AACCGTTGTATACACGCGCGCGCGCATCCTTTATCGCTGTTTCGCGCATTAATCTAC 786
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 245 LysAlaLeuAenThrThrAlaAlaGlyAspThrPheValGlyGlyPheAlaAlaLeu 264
Qy 787 GTCCAGAGCGGGATGTGAAGCGCCCATGAAAAAGCCGCTCTTTGCGCGCTTTCAGC 846
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 265 AlaArgGlyLeuAspGluAlaAlaAlaIleArgPheGlyGlnAlaAlaAlaIleSer 284
Qy 847 GTCCCGGGAAGGACCCCATCTCTTATCCAAAGCATTTGACCAATTTAATGAGTATCTT 906
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 285 ValThrArgLeuGlyAlaGlnThrSerIleProSerArgGluValGluArgAlaLeu 304

RESULT 5
T20529
hypothetical protein F07A11.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T20529
R:Palmer, S.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z19287
A:Accession: T20529
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-343 <WIL>
A:Cross-references: UNIPROT:Q19133; UNIPARC:UPI000008223A; EMBL:Z66511; PIDN:CAA91318.1;
A:Experimental source: clone F07A11
C:Genetics:
A:Gene: CESP:F07A11.5
A:Map position: 2
A:Introns: 19/1; 49/2; 136/1; 221/3; 259/1; 284/3
C:Superfamily: ribokinase
```

```
Alignment Scores:
Pred. No.: 1,07e-37 Length: 343
Score: 523.00 Matches: 114
Percent Similarity: 56.54% Conservative: 59
Best Local Similarity: 37.25% Mismatches: 127
Query Match: 32.89% Indels: 6
DB: 2 Gaps: 2
```

```
US-10-049-750-11 (1-921) x T20529 (1-343)
Qy 7 ATCCGGCTTATTGCTCTTAACATCGTGGACCTTATCATTACACCAACCATGATGCCCAA 66
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 37 IleValValPheGlySerIleValGlnAspLeuValSerTyrThrAspSerPheProArg 56
Qy 67 GAAGGGGAAACTCTCGGAAGCGCGCGTAAATCGCTCGCGCGGAAAAAGGCGCGAAC 126
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 57 ProGlyGluSerValArgGlyHisAsnPhelysLeuGlySerGlyGlyLysGlyAlaAsn 76
Qy 127 CAGCGCGTGGCGCGCTCAAGCTCAATTCAAAAAGTATTGATGTTGACCAAAAGTGGCGAC 186
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 77 GlnAlaValAlaAlaAlaArgLeuGlyLeuAsnValSerMetIleGlyMetValGlyGlu 96
Qy 187 GATATTTTTCGCGCAACACACCATTCGTAATCTCAATCTCGAATCTCGGGGATCAATACGACGTAT 246
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 97 AspMetPheGlyAspSerAsnIleLysAspLeuSerSerAsnGlyValAspThrSerCys 116
Qy 247 GTAGAAAAAGTACCGTGTACCAGCGCGGTAGCGCGATTTTCGTCAACGCCAACTCC 306
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 117 ValGlyArgThrLysLysThrHisThrAlaThrAlaThrIleThrValAsnLysGluGly 136
Qy 307 AGCAACAGCATTCGTGATCATCAAAAGCGCTAACAAAGTTTCTCTCGCGGAAGATATCGAT 366
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 137 GluAenAenIleValValThrLeuGlyAlaAenLeuGluMetSerProGluIleAlaAen 156
Qy 367 CGCGCGCGGAGATTAAAAAATGCCAGCTTATTGTTCTGCACTGGAAGTTCAGCTT 426
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 157 AlaAenSerSerIleIleAlaGlySerLysMetValIleCysGlnGlyGluIleAenGlu 176
Qy 427 GAAACGGTTTATCAGCAATAGAAATTTGGCAAGAAACACACGGGATTGAAAGTGTATTAAAC 486
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 177 LysGlyAenArgArgAlaPheGluIleAlaArgSerHisGlyValThrThrPheLeuAen 196
Qy 487 CTCTGCGCAGCATTAACGGGAATTAGATATGCTTATGCTGCTGTAATGCGATTTCTTTGTA 546
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 197 ProAlaProGlyAspProAenMetAspLysThrIleLeuGluLeuValAspIleCys 216
Qy 547 CCTAATGAACCGAGCTGGAATATTAAACGGTATGCCAGTATGCCAGTATGACCATATT 606
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 217 ThrAenGluAenGluAlaGluPheIleThrGlyIleAlaGlnAenAenValLysAspAla 236
Qy 607 CGCGCAGCGCAGCTGCTGCTAGATAAAGCGCTGAACAATATTATTGTCACCATGGC 666
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 237 GluLysAlaAlaArgAlaMetLeuLysMetGlyProGlnHisAlaIleIleThrLeuGly 256
Qy 667 GAGAAAGCGCGCTGTGGATGACCGCT-----GACCAGGAAGTCCATGTTTCGCGCGTTT 720
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 257 GlyLysGlyValLeuLeuAlaSerLysGlyValAspAspValGluHisThrAlaValIle 276
Qy 721 AGAGTGAACGCTGTTGATACACGCGCGCGCGCATGCTTTATCGCTGTTTCGCGCAT 780
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 277 LysValAspAlaValAspThrThrGlyAlaGlyAspCysPheCysGlySerLeuAlaAla 296
Qy 781 TACTACGTCCAGAGCGGGATGTGAAGCGCCCATGAAAAAGCGCTCTCTTTGCGCGCT 840
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 297 HisLeuValAlaAlaGlyHisProIleSerAlaSerIleArgSerAlaAlaAenLeuAlaAla 316
Qy 841 TTACAGCTTCACCGGGAAGGACCCCAATCTCTTTATCCAAAGCATT-----GAG 888
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 317 LeuSerValThrArgHisGlyThrGlnSerSerTyrTrpLysLeuAenGluIleArgGln 336
Qy 889 CAATTTAATGATATCTT 906
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 337 GlnTyrProGluPheLeu 342

RESULT 6
G89790
hypothetical protein rbsK [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: G89790
```

R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogud
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A>Title: Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*.

A:Reference number: A89758; PMID:21311952; PMID:11418146

A:Accession: G89790

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-304 <KUR>

A:Cross-references: UNIPROT:Q99WV7; UNIPARC:UPI00000CA992; GB:BA000018; PID:gl3700184; E

A:Experimental source: strain N315

C:Genetics:

A:Gene: rbsK

C:Superfamily: ribokinase

Alignment Scores:
Pred. No.: 5.6e-34 Length: 304
Score: 480.50 Matches: 105
Percent Similarity: 55.45% Conservative: 63
Best Local Similarity: 34.65% Mismatches: 126
Query Match: 30.22% Indels: 9
Gaps: 4

US-10-049-750-11 (1-921) x G89790 (1-304)

Qy 7 ATCCGGTTATTGGCTTAACATGTCGGACCTTATCACCTACCAACACAGATGCCCAA 66
Db 5 ValValIleLeuGlySerThrAsnValAspGlnPheLeuThrValGluArgTyrAlaGln 24
Qy 67 GAAGGGGAACTCTGGAAAGCG---CCGGCGTTTAAATCGCGCGGCGGAAAGGGCG 123
Db 25 ProGlyGluThrLeuHisValGluGlnAlaGlnLysAlaPheGlyGlyLysGlyAla 44
Qy 124 AACGAGCGCGCGCGCGCTTAAGCTCAATCAAAAGTATTGATGTTTACCAAGTGGC 183
Db 45 AsnGlnAlaIleAlaThrAlaArgMetGlnAlaAspThrPheIleThrLysIleGly 64
Qy 184 GACGATATTTTGGCGACACACCATTCGTATCTCGAATCTGGGGATCAATACGACG 243
Db 65 ThrAspGlyValAlaAspPheIleLeuGluAspPheLysAlaAlaHisIleAspThrSer 84
Qy 244 TATGTAGAAAAGTAGTACCGGTGTACACGACGCGGTAGCGCGGATTTTCGTCAACGCCAAC 303
Db 85 TyrIleIleLysThrThrGluAlaLysThrGlyGlnAlaPheIleThrValAsnAlaGlu 104
Qy 304 TCCGACACAGCATTTCTGATCATCAAGCGGTAAAGTTTCTCTCCGCGGAAGATATC 363
Db 105 GlyGlnAsnThrIleTyrValTyrGlyGlyAlaAsnMetThrMetThrProGluAspVal 124
Qy 364 GATCCGCGCGGCGAAGATTAAATAATGCCAGCTTATTGTTCTGCAACTGGAAGTTCAG 423
Db 125 IleAsnAlaLysAspAlaIleIleAsnAlaAspPheValValAlaGlnLeuGluValPro 144
Qy 424 CTTGAACCGTTTATCAGCAATAGAAATTTGGCAAGAAACACGGGATGAAGTGTATTA 483
Db 145 IleProAlaIleIleSerAlaPheGluIleAlaLysAlaHisGlyValThrThrValLeu 164
Qy 484 AACCTCGCGCCAGCATTAAGGAATAGATATGCTTATGCTGCTGAATGCGATTTCTTT 543
Db 165 AsnProAlaProAla---LysAlaLeuProAsnGluLeuLeuSerLeuIleAspIleIle 183
Qy 544 GTACCTAATGAACGAGTGAATATTAATTAACCGGTATGCCAGTGGATACCTTACCAT 603
Db 184 ValProAsnGluThrGluAlaGluLeuSerGlyIleLysValThrAsnGluGlnSer 203
Qy 604 ATTCCGCGCGCGCGCTTCGCTGTAGATAAAGGCGCTCAACATATATTATTCACCATG 663
Db 204 MetLysAspAsnAlaAsnTyrPheLeuSerLeuGlyIleLysThrValLeuIleThrLeu 223
Qy 664 GCGGAGAAGCGCGCTGTGGATGACGCGTGACCGAGTACCAAGAGTCCATGTTCCGCGGTTAGA 723
Db 224 GlyLysGlnGlyThrTyrPheAlaThrLysAsnGlnSerGlnHisIleGluAlaTyrLys 243

Qy 724 GTGAACGCTGTTGATACACAGCGCGCGGCATCCCTTTATCGCTGTTTCGCGCATTAC 783
Db 244 ValAsnAlaIleAspThrThrAlaAlaGlyAspThrPheIleGlyAlaPheValSerArg 263
Qy 784 TACGTCCAGACGCGGGATGTGAAGCGCCCATGAAAAAAGCGCTCTCTTT----- 834
Db 264 LeuAsnLysSerGlnAsp-----AsnLeuAlaAspAlaIleAspPheGlyAsnLys 280
Qy 835 ---GCCGCTTTTCACGCTCACCGGAAAGGACCCCAATCCTTTATCCAGCATTCAGCAA 891
Db 281 AlaSerSerLeuThrValGlnLysHisGlyAlaGlnAlaSerIleProLeuLeuGluGlu 300
Qy 892 TTTAATGAG 900
Db 301 ValAsnGln 303

RESULT 7

D69690

ribokinase (EC 2.7.1.15) - *Bacillus subtilis*

C:Species: *Bacillus subtilis*

C>Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004

C:Accession: D69690; I40463; S42711

R:Kunst, F.; Ogasawara, N.; Mozer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle

teck, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.

A:Authors: Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidis, A.; Lardinois,

Y. M.; Ogawa, K.; Ogawa, B.; Oudega, B.; Park, S.H.; Levine, A.; Liu, H.; Maguda, S.; Maue

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, R.; Sekiguchi, J.; Sekowska, A.; Ser

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyana,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yanane, K.; Yasumoto, K.; Yata, K.; Yoshida,

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A. *Bacillus subtilis*.

A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; PMID:98044033; PMID:9384377

A:Accession: D69690

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-293 <KUN>

A:Cross-references: UNIPROT:P36945; UNIPARC:UPI000006084E; GB:Z99122; GB:AL009126; NID:5

A:Experimental source: strain 168

R:Woodson, K.; Devine, K.M.

Microbiology 140, 1829-1838, 1994

A>Title: Analysis of a ribose transport operon from *Bacillus subtilis*.

A:Reference number: I40462; MUID:95005437; PMID:7921236

A:Accession: I40463

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-283 'DKK' <RES>

A:Cross-references: UNIPARC:UPI0000168934; EMBL:Z25798; NID:9397493; PIDN:CAA81049.1; PI

C:Comment: This enzyme catalyzes the phosphorylation of ribose to ribose 5-phosphate, t

C:Genetics:

A:Gene: rbsK

C:Superfamily: ribokinase

C:Keywords: phosphotransferase; ribose metabolism

Alignment Scores:
Pred. No.: 1.87e-33 Length: 293
Score: 474.50 Matches: 114
Percent Similarity: 55.30% Conservative: 53
Best Local Similarity: 37.75% Mismatches: 124
Query Match: 29.84% Indels: 11
Gaps: 4

US-10-049-750-11 (1-921) x D69690 (1-293)

Qy 4 GATATCGCGTTATTGGCTTAACATGTTGGACCTTATCACCTACCAACACGATGCC 63
Db 3 AnIleCysValIleGlySerCysSerMetAspLeuValThrSerAspLysArgPro 22


```
QY 784 TACGTCACAGCGGGATGTGGAAGCCCGCATGAAAGAGCGTCCTCTTTGCGCGCTTTC 843
Db 265 ILeuGluGluThrProLeuProGluAlaIleArgPheAlaHisAlaAlaAlaIle 284
QY 844 AGCGTCACCGGGAAGGACCAATCCTCTTATCCAGCATTTGACCAATTTAATGAGTAT 903
Db 285 AlaValThrArgLysGlyAlaGlnProSerValProTrpArgGluGluIleAlaPhe 304
QY 904 CTT 906
Db 305 Leu 305

RESULT 9
T39594
ribokinase - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T39594
R:Lyne, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z21866
A:Accession: T39594
A>Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-318 <LYN>
A:Cross-references: UNIPROT:O60116; UNIPARC:UPI0000133387; EMBL:AL023554; PIDN:CAA19022.
A:Experimental source: strain 972h-; cosmid c16G5
C:Genetics:
A:Gene: SPDB:SPBC16G5.02c
A:Map position: 2
C:Superfamily: ribokinase

Alignment Scores:
Pred. No.: 6.39e-33 Length: 318
Score: 468.50 Matches: 115
Percent Similarity: 55.06% Conservative: 59
Best Local Similarity: 36.39% Mismatches: 131
Query Match: 29.47% Indels: 11
DB: 2 Gaps: 6

US-10-049-750-11 (1-921) x T39594 (1-318)
QY 1 ATGGATATCGCGTTATTGGCTCTTAACATGTTGGAGCTTATCATCTACCAACCAACAGATG 60
Db 2 IleAsnIleValValLeuGlySerMetAsnThrAspLeuValMetArgThrIleCys 21
QY 61 CCCAAGAAAGGGAACTCTG-----GAAGCGCGCGGCTTTAAATCGCGTGGCGGA 114
Db 22 ProSerGlyGluThrIleHisGlyGluProAspGlyPheSerThrGlyAsnGlyGly 41
QY 115 AAAGGGCGAACAGCGCGTGGCGCGCTAAGCTC-----AATTCAAAGTA 162
Db 42 LysGlyAlaAsnGlnAlaValAlaValAlaArgLeuSerAsnProAlaAspThrIleVal 61
QY 163 TTGATGTTCCACCAAGTGGCGGACGATATTTTGGCGACACCAACCATTCGTAATCTCGAA 222
Db 62 SerMetLeuGlyCysValGlyAspAspAlaPheGlyValGluMetLeuSerGlyLeuLys 81
QY 223 TCCTGGGGATCAATACGACGTGTGTAGAAAAGTACCGTGTACGAGCGCGGTAGG 282
Db 82 LysAspGlyValAsnValAspAsnValLysIleGluAsnLysSerThrGlyValAla 101
QY 283 CCGATTTTCGTAACGCCAACCTCCAGCAACAGATTCTCATCAAGCGCTCAACAG 342
Db 102 MetIleValGluGluThrGlyGluAsnArgIleLeuLeuSerGluGlyAlaAsnGly 121
QY 343 TTCTCTCGCGGAAGATATCGATCGCGCGGGAAGATTAAAAAATGCCAGCTTAT 402
Db 122 AsnValAspThrAlaPheValLysAlaMetGluGlnArgIleSerThrCysAsnLeuLeu 141
QY 403 GTTCTGCAACTGGAAGTTCAGCTTGAACCGTTTATCAGCAATAGAAATTTGGCAGAAA 462
Db 142 IleMetGlnLeuGluIleProLeuGluAlaValGluIleAlaLeuGlnIleAlaHisLys 161
```

```
QY 463 CACGGGATTTGAAGTGTATTATTAACCTCGCGCCAGCATTTACGGGAATTAGATATGCTTAT 522
Db 162 HisGlyValAspValLeuMetAsnProAlaProAlaIle---ProLeuSerHisAspMet 180
QY 523 GCCTGTAATCGGATTTCTTGTACCTAATCAAAACCGAGCTGGAAATATTAAACCGGTATG 582
Db 181 IleSerTyrCysAlaIleValProAsnGluHisGluAlaAlaIleLeuLeuAsnGln 200
QY 583 CCAGTGGAT-----ACCTATGACCATATTCCGCGAGCGGCACGCTTCCTGGTAGATAAA 636
Db 201 AlaAspSerProAlaThrLeuGluAsnValAspAlaIleValAspLeuLeuSerPhe 220
QY 637 GGGCTGAAACAT---ATTATTGTACCATGGCGGAGAAAGCGCGCTGTGGATGACGCGT 693
Db 221 GlyValArgLysAlaValIleThrLeuGlySerGlnGlyAlaIleTyrTyrLysSerAla 240
QY 694 GACCAGGAA---GTCATGTTCCGCGTGTAGAGTGAACGCTGTGTGATACAGCGCGCG 750
Db 241 AsnGlyGluSerAlaLeuValSerAlaCysLysValLysAlaValAspThrThrAlaAla 260
QY 751 GCGCATGCTTTTATCGGCTGTTTCGCGCATTTACTACGTCAGAGCGGGATGTGGAAGCC 810
Db 261 GlyAspThrPheIleGlyAlaPheSerAsnSerIleAlaHisGlyGlnProLeuLysAsp 280
QY 811 GCCATGAAAAAGCGCTCTCTTTCGCCCTTTCACGCTCACCGGAAAGGACCCCAATCC 870
Db 281 SerLeuGluPheAlaAlaLysCysSerAlaIleThrValGlnArgLysGlyAlaAlaSer 300
QY 871 TCTTATCCAGCATTTGACCAATTTAATGAGTATCTTCTGTTGAACGAA 918
Db 301 SerIleProSerLeuLeuGluValAspGlySerPheAsnLeuLysLys 316
```

```
RESULT 10
A10001
ribokinase (EC 2.7.1.15) [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C:Accession: A10001
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: A10001
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-308 <KUR>
A:Cross-references: UNIPROT:Q82JS8; UNIPARC:UPI00000DCD63; GB:AL590842; PIDN:CAC88875.1
C:Genetics:
C:Gene: rbsK
C:Superfamily: ribokinase
C:Keywords: phosphotransferase
```

```
Alignment Scores:
Pred. No.: 1.74e-32 Length: 308
Score: 463.50 Matches: 109
Percent Similarity: 55.48% Conservative: 58
Best Local Similarity: 36.21% Mismatches: 131
Query Match: 29.15% Indels: 3
DB: 2 Gaps: 3
```

US-10-049-750-11 (1-921) x A10001 (1-308)

```
QY 7 ATCGCGGTATTGGCTCTTAACATGTTGGACCTTATCATCTACCAACCAACAGATGCCCAA 66
Db 6 LeuValValLeuGlySerIleAsnAlaAspHisIleLeuAsnIleGluGlnPheProArg 25
QY 67 GAAGGGGAACCTCTGGAAGCGCGCTTTAAATTCGGTTCGGCGGGAAGGGGCGAAG 126
Db 26 ProGlyGluThrValIleGlyGlnGlnTyrAsnValAlaPheGlyLysGlyAlaAsn 45
```

```
Qy 127 CAGCCGTCGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAAAGTGGCGAC 186
Db 127 CAGCCGTCGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAAAGTGGCGAC 186
46 GlnAlaValAlaAlaGlyArgSerGlyAlaAspIleAlaPheIleAlaCysValGlyAsp 65
187 GATATTTTGGCGCAACACCATTCGTATCTCGAATCTCGGGGATCAATACGAGTAT 246
Db 187 GATATTTTGGCGCAACACCATTCGTATCTCGAATCTCGGGGATCAATACGAGTAT 246
66 AspAspIleGlyGluArgValArgGlnGlnLeuThrAlaAspLysIleAspThrGlnPro 85
247 GTAGAAAAGTACCGTGTACAGCAGCGCGCTAGCGGATTTTCGTCACCGCAACTCC 306
Db 247 GTAGAAAAGTACCGTGTACAGCAGCGCGCTAGCGGATTTTCGTCACCGCAACTCC 306
86 IleGluAlaIleLysGlyAlaThrThrGlyValAlaLeuIlePheValAsnSerAspGly 105
307 AGCAACAGCATTCGATCATCAAGAGCGCTAACAAAGTTCTCTCGCGGGAAGATATCGAT 366
Db 307 AGCAACAGCATTCGATCATCAAGAGCGCTAACAAAGTTCTCTCGCGGGAAGATATCGAT 366
106 GluAsnValIleGlyIleAsnAlaGlyAlaAsnSerAlaValThrProGluTyLeuArg 125
367 CGCGCGCGGAGAGATTAAAAAATGCGCAGCTTATTGTTCTGCAACTGGAAGTTTCAGCTT 426
Db 367 CGCGCGCGGAGAGATTAAAAAATGCGCAGCTTATTGTTCTGCAACTGGAAGTTTCAGCTT 426
126 ArgTyGlnGlnGlnValIleAspAlaAspAlaLeuLeuMetGlnLeuGluSerProLeu 145
427 GAAACGGTTTATCAGCAGTATGAAATTTGGCAAGAAACACGGGATGAAAGTCTTAAAC 486
Db 427 GAAACGGTTTATCAGCAGTATGAAATTTGGCAAGAAACACGGGATGAAAGTCTTAAAC 486
146 AspThrValIleAlaAlaLysLeuAlaLysGlnHisGlnThrGlnValIleLeuAsn 165
487 CTGCGCGCAGCATTCAGCGGAATTAGATATGCTTATGCTCTGTAATGCGATTTCTTGTGA 546
Db 487 CTGCGCGCAGCATTCAGCGGAATTAGATATGCTTATGCTCTGTAATGCGATTTCTTGTGA 546
166 ProAlaProAla--ArgLysLeuProAspGluLeuLeuThrLeuValAspMetIleThr 184
547 CCTAATGAACCGAGCTGGAATATTAAACCGTATGCGAGTGCGATACCTATGACCATATT 606
Db 547 CCTAATGAACCGAGCTGGAATATTAAACCGTATGCGAGTGCGATACCTATGACCATATT 606
185 ProAsnGluThrGluAlaGluArgLeuThrGlyIleHisIleGluGlnAspAspAla 204
607 CGCGCAGCGGACGTTCTGCTGATGAAGAGGCTGAACATATATTGTCACCATGGC 666
Db 607 CGCGCAGCGGACGTTCTGCTGATGAAGAGGCTGAACATATATTGTCACCATGGC 666
205 AlalysAlaAlaGlnLeuHisAspLysGlyIleAlaThrValIleIleThrLeuGly 224
723 GAGAAAGCGCGCTGTGATGATGACGCGTACAGCAAGTCCAT--GTTCCGCGCTTTAGA 723
Db 723 GAGAAAGCGCGCTGTGATGATGACGCGTACAGCAAGTCCAT--GTTCCGCGCTTTAGA 723
225 SerArgGly--ValTrpLeuSerGluGlnGlyLysGlyLysLeuValAlaGlyPheLys 243
724 GTGAACCGTGTGTATACAGCGCGCGCGGATGCTTTTATCGGCTGTTTCGCGCATTAC 783
Db 724 GTGAACCGTGTGTATACAGCGCGCGCGGATGCTTTTATCGGCTGTTTCGCGCATTAC 783
244 ValAsnAlaValAspThrIleAlaAlaGlyAspThrPheAsnGlyAlaLeuLeuThrAla 263
843 TACGTCACAGCGGGGATGTGAAGCCCATGAAAGAGCGCTCTTTCGCGCTTTC 843
Db 843 TACGTCACAGCGGGGATGTGAAGCCCATGAAAGAGCGCTCTTTCGCGCTTTC 843
264 LeuLeuGluGlyGlnSerMetAspValAlaValArgPheAlaHisAlaAlaAlaIle 283
904 CTT 906
Db 904 CTT 906
304 Leu 304
RESULT 11
H84115
ribokinasase rbsk [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C:Accession: H84115
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: H84115
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-294 <STO>
A:Cross-references: UNIPROT:Q9K6K1; UNIPARC:UPI00001333E4; GB:AP001519; GB:BA000004; NID
A:Experimental source: strain C-125
C:Genetics:
```

```
A:Gene: rbsk
C:Superfamily: ribokinasase
Alignment Scores:
Pred. No.: 3.5e-32 Length: 294
Score: 460.00 Matches: 113
Percent Similarity: 55.81% Conservative: 55
Best Local Similarity: 37.54% Mismatches: 117
Query Match: 28.93% Indels: 16
DB: 2 Gaps: 5
US-10-049-750-11 (1-921) x H84115 (1-294)
Qy 4 GATATCGCGGTTATTGGCTCTTAACATGGTGGGACCTTATCACCTACACCAACGAGTGGCC 63
Db 4 GATATCGCGGTTATTGGCTCTTAACATGGTGGGACCTTATCACCTACACCAACGAGTGGCC 63
5 AsnIleThrValValGlySerIleAsnMetAspMetValThrIleThrAspValValPro 24
64 AAGAAGGGGAACTCTGGAAGCGCGCGTTTAAATCGGCTCGCGGCGGAAAGGGGCG 123
Db 64 AAGAAGGGGAACTCTGGAAGCGCGCGTTTAAATCGGCTCGCGGCGGAAAGGGGCG 123
25 ValGlnGlyGluThrValLeuGlyLysAspPheArgThrValProGlyGlyLysGlyAla 44
124 AACGAGCGCGTGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAAAGTGGC 183
Db 124 AACGAGCGCGTGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAAAGTGGC 183
45 AsnGlnAlaValAlaAlaAlaArgLeuGlyAlaAsnValArgMetIleGlyArgValGly 64
184 GACGATATTTTTCGCGCAACACCATTCGTAATCTCGAATCTCGGGGATCAATACGACG 243
Db 184 GACGATATTTTTCGCGCAACACCATTCGTAATCTCGAATCTCGGGGATCAATACGACG 243
65 AspAspProPheGlyHisValLeuThrGluAsnLeuAlaLysGlyIleIleThrAsp 84
244 TATGTAGAAAAGTACCGCTGTACAGCAGCGCGCTAGCGCGGATTTTCGTCACCGCCAC 303
Db 244 TATGTAGAAAAGTACCGCTGTACAGCAGCGCGCTAGCGCGGATTTTCGTCACCGCCAC 303
85 SerValLysProValThrAspCysThrSerGlyValAlaThrIleLeuLeu--SerAsp 103
304 TCCAGCACACGATTCGATCATCAAGGGCGTCAACAGTTTCTCTCGCGGAAAGATATC 363
Db 304 TCCAGCACACGATTCGATCATCAAGGGCGTCAACAGTTTCTCTCGCGGAAAGATATC 363
104 ArgAspAsnArgIleIleValThrLysGlyAlaAsnGluHisValThrProAspTyVal 123
364 GATCGCGCGCGGAGAGTTTAAAAAATGCGCAGCTTATTGTTCTGCAACTGGAAGTTTCAG 423
Db 364 GATCGCGCGCGGAGAGTTTAAAAAATGCGCAGCTTATTGTTCTGCAACTGGAAGTTTCAG 423
124 AlaAlaPheGluGlnGluLeuAlaAlaSerAspValValLeuLeuGlnLeuGluIlePro 143
424 CTTGAAACCGTGTATACAGCAATAGAAATTTGGCAAGAAACACCGGATTTGAAGTGTATTA 483
Db 424 CTTGAAACCGTGTATACAGCAATAGAAATTTGGCAAGAAACACCGGATTTGAAGTGTATTA 483
144 LeuGluThrValAlaTyValLeuGluPheCysAlaLysHisHisValThrValLeu 163
484 AACCTCGCGCAGCATTTACGGGAATTAGATATGCTTATGCTCTGTAATGCGATTTCTT 543
Db 484 AACCTCGCGCAGCATTTACGGGAATTAGATATGCTTATGCTCTGTAATGCGATTTCTT 543
164 AsnProAlaProAlaGlnLysLeuProAspAlaAlaIleThr--AspAlaThrTyIle 182
544 GTACCTAATGAACCGAGCTGGAATATTAAACCGTATGCGAGTGATACCTATGACCAT 603
Db 544 GTACCTAATGAACCGAGCTGGAATATTAAACCGTATGCGAGTGATACCTATGACCAT 603
183 SerProAsnGluAsnGluCysLeuGlnLeuPheGlyAspGluProAspAla-----Asn 200
604 ATTCGCGCAGCGCACGCTTCGCTGCTAGATAAAGGCTGAACAATATTATTGTCACCATG 663
Db 604 ATTCGCGCAGCGCACGCTTCGCTGCTAGATAAAGGCTGAACAATATTATTGTCACCATG 663
201 LeuArg-----GlnLysLeuIleMetThrLys 209
664 GCGCAGAAAGCGCGCTGTGGATACGCGTGACCGAGGAGTCCATGTTCCGCGCTTTAGA 723
Db 664 GCGCAGAAAGCGCGCTGTGGATACGCGTGACCGAGGAGTCCATGTTCCGCGCTTTAGA 723
210 GlyAlaAspGlyValGlnPheTyArgGluAsnAspGluGlnValGlnValGlnSerPheArg 229
724 GTGAACCGTGTGTATACAGCGCGCGCGGATGCTTTTATCGGCTGTTTCGCGCATTAC 783
Db 724 GTGAACCGTGTGTATACAGCGCGCGCGGATGCTTTTATCGGCTGTTTCGCGCATTAC 783
230 ValGluProValAspThrThrGlyAlaGlyAspThrPheAsnGlyAlaPheAla---Val 248
784 TACGTCACAGCGGGGATGTGAAGCCCATGAAAGAGCGCTCTTTCGCGCTTTC 843
Db 784 TACGTCACAGCGGGGATGTGAAGCCCATGAAAGAGCGCTCTTTCGCGCTTTC 843
249 AlaLeuGlyGlyThrValLysGluAlaValArgPheAlaAsnAlaAlaAlaLeu 268
844 AGCGTCACCGGGAAGGACCCCAATCTCTTATCCAGCATTTGAGCAATTTAATGAGTAT 903
Db 844 AGCGTCACCGGGAAGGACCCCAATCTCTTATCCAGCATTTGAGCAATTTAATGAGTAT 903
269 SerValGlnSerPheGlyAlaGlnGlyGlyMetProThrLysAlaGlnValGlnSerPhe 288
```

```
Qy 904 CTT 906
Db 289 Leu 289

RESULT 12
Ti3380
ribokinase homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:115C2.1
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C:Accession: T13380
R:Salles, C.; Valenti, P.; Darlamiteou, A.; Henderson, N.; Campbell, L.; Glover, D.
submitted to the EMBL Data Library, May 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17665
A:Accession: T13380
A>Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-304 <CAT>
A:Cross-references: UNIPROT:O77425; UNIPARC:UPI000008263F; EMBL:AL031581; PIDN:CAA20884.
C:Genetics:
A:Cross-references: FlyBase:FBgn0020381
A:Map position: X
A:Introns: 18/2; 201/2
A:Note: EG:115C2.1
C:Superfamily: ribokinase

Alignment Scores:
Pred. No.: 3,53e-32 Length: 304
Score: 460.00 Matches: 109
Percent Similarity: 56.48% Conservative: 61
Best Local Similarity: 36.21% Mismatches: 119
Query Match: 28.93% Indels: 12
Gaps: 6

US-10-049-750-11 (1-921) x T13380 (1-304)
Qy 4 GATATCGCGTATTGGTCTTAACATGTTGGACCTTATCACCTACACCAACAGATGCC 63
Db 5 GluValLeuValPheGlySerAlaIleAaspPheIleSerTyThrArgLeuPro 24
Qy 64' AAAGAGGGGAACTCTGGAAGCGCGGCTTTAAATCGGTCGGCGGCGGAAAGGGGG 123
Db 25 LysAlaGlyGluThrLeuHisGlyHisArgPheGlnIleGlyTyThrGlyLysGlyAla 44
Qy 124 AACAGCGCGTGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGGC 183
Db 45 AsnGlnCysValAlaAlaAlaArgGlnGlySerA-gThrAlaLeuValAlaLysLeuGly 64
Qy 184 GACGATATTTTGGCGACACACCATTCGTATCTCGAATCTCGGGGATCAATACGAG 243
Db 65 AlaAspThrPheGlySerAspTyThrLeuArgHisLeuArgGluGluArgValAsnValAsn 84
Qy 244 TATGTAGAAAAGTACCGGTGTACACGACGCGCGTAGCGCGATTTTGTCAACGCCAAC 303
Db 85 HisValGluGlnLeuAlaGluThrGlyValAlaGlnIleAlaValSerAspGly 104
Qy 304 TCACGACAGCATTCGTGATCATCAAGCGCGTCAAGTTTCTCTCGCGGGAAGATATC 363
Db 105 GlyGluAsnAsnIleIleValValGlyAlaAsnAsnArgLeuSerSerCysAspVal 124
Qy 364 GATCGCGCGGGAAGATTTAAAAATGCCAGCTTATTGTTCTGCAACTGGAAGTTCAG 423
Db 125 SerSerAlaLysAlaLeuPheGlnGluAlaLysValLeuValCysGlnLeuGluThrPro 144
Qy 424 CTTGAAACGGTTTATCACGCAATAGAAATTTGGCAAGAAACACGCGGATGAAGTGTATTA 483
Db 145 ValGluAlaThrLeuThrAlaLeuArgAlaPheArg-----GlyValSerile---Val 161
Qy 484 AACCCCTGCCAGCATTCGGGAATTAGATATGCTTATGCTGCTGTAATGCAATTCCTTT 543
Db 162 AsnAlaAlaProAlaMetAlaAspThrProGluLeuLeuGlnLeuAlaSerIlePhe 181
```

```
Qy 544 GTACCTAATGAAACCGAGCTGGAATATTAAACCGGTATGCCA---GTGATACCTATGAC 600
Db 182 CysValAsnGluSerGluAlaAlaLeuMetThrGlnMetProAspIleGlyAsnIleGlu 201
Qy 601 CATATTGCGCGCAGCGCACGCTTCGTGTAGATAAAGGCTGGAACAATATTATTGTACCC 660
Db 202 HisAlaGluAspAlaValGlyLysLeuIleAlaAlaGlyAlaAsnThrValIleIleThr 221
Qy 661 ATGGCGGAGAAAGCGCGCTGTGGATGACCGCTGACACGAGAA-----GTCCATGTT 711
Db 222 LeuGlyLysLeuGlyAlaValPheGlySerAlaAspSerLysGlyValCysGlnHisVal 241
Qy 712 CCGCGCTTACAGTG-----AACGCTGTGTATACCGCGCGCGGCGATGCCCTTT 762
Db 242 AlaAlaProSerValProGluLysValValAspThrThrGlyAlaGlyAspAlaPhe 261
Qy 763 ATCGGCTGTTTCGGCGCATTTACTAGCTCCAG-----AGCGGGGATGTGGAAGCCCGCATG 816
Db 262 IleGlyAlaLeuAlaHisAsnLeuAlaArgHisProThrArgLysLeuGluGluHisIle 281
Qy 817 AAAAAAGCGTCTCTTTGCGCTTACGCTACCGGGAAGGACCCCAATCTCTTAT 876
Db 282 AlaAlaLysAlaValAlaSerGlnSerValGlnLeuProGlyThrGlnSerSerPhe 301
Qy 877 CCA 879
Db 302 Pro 302

RESULT 13
B64073
ribokinase (EC 2.7.1.15) - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004
C:Accession: B64073
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, J.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.W.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: B64073
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-306 <TIGR>
A:Cross-references: UNIPROT:P44331; UNIPARC:UPI00001333E5; GB:U32732; GB:L42023; NID:gl
C:Superfamily: ribokinase
C:Keywords: phosphotransferase; ribose metabolism

Alignment Scores:
Pred. No.: 2,41e-31 Length: 306
Score: 450.50 Matches: 108
Percent Similarity: 53.33% Conservative: 52
Best Local Similarity: 36.00% Mismatches: 139
Query Match: 28.33% Indels: 1
Gaps: 2

US-10-049-750-11 (1-921) x B64073 (1-306)
Qy 7 ATCGCGTATTGGTCTTAACATGTTGGACCTTATCACCTACCAACCCAGATGCCAAA 66
Db 5 LeuThrValLeuGlySerIleAsnAlaAspHisValIleSerValProTyThrPheThrLys 24
Qy 67 GAAGGGGAAACTCTGGAAGCGCGCTTTAAATCGGCTGCGGGGAAAAGGGCGCAAC 126
Db 25 ProGlyGluThrLeuThrGlyGlnAsnTyThrGlnIleAlaTyThrGlyLysGlyAlaAsn 44
Qy 127 CAGGCGCGCGCGCTCAAGCTCAATCAAAGTATTGATGTTGACCAAGTGGCGCAC 186
Db 45 GlnAlaValAlaAlaAlaArgLeuGlyAlaLysValAlaPheIleSerCysIleGlySer 64
Qy 187 GATATTTTGGCGCACACACCATTCGTAATCTCGAATCTCGGGGATCAATACGAGTAT 246
```


Qy 847 GTCCACGGGAAGGCCCAATCCTTTATTCGACGATTGAGCAATTAATGAGTATCTT 906
||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 286 ValThrArgLysGlyAlaGlnProSerValProTrpArgGluGluLeuAspAlapheLeu 305
||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15

F91215 ribokinase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: F91215
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawaka, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen-
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F91215
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-309 <HAY>
A:Cross-references: UNIPROT:P05054; UNIPARC:UPI000003EB7F; GB:BAO000007; PIDN:BAB38117.1/
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECS4694
C:Superfamily: ribokinase

Search completed: December 23, 2005, 22:46:38
Job time : 48.015 secs

THIS PAGE BLANK (USE)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 23, 2005, 22:15:36 ; Search time 248.618 Seconds

(without alignments)
5227.239 Million cell updates/sec

Title: US-10-049-750-11

Perfect score: 1590

Sequence: 1 atggatcggtattgg.....atcttcgtgaacgaataa 921

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_gpool_p/US10049750/runat_23122005_113552_6292/app_query.fasta_1.1742
-DB=UniProt -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HRAPSZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10049750@cgn_1_1_614@runat_23122005_113552_6292 -NCPU=6 -ICPU=3
-NO MMAP -LARGSEQURY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt 05.80:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1570	98.7	306	Q7CPF5_SALTY	Q7cpf5 salmonella
2	1570	98.7	306	Q8XGW5_SALTI	Q8xgw5 salmonella
3	1570	98.7	306	Q5PN81_SALPA	Q5pn81 salmonella
4	1563	98.3	306	Q57143_SALCH	Q57143 salmonella
5	1519	95.5	306	Q67FX7_ESCHRIC	Q67fx7 escherichia
6	1519	95.5	306	Q8FKR4_ECOLI	Q8fkr4 escherichia
7	1519	95.5	315	Q70712_ECOLI	Q70712 escherichia
8	1435.5	90.3	314	Q9A119_ECOLI	Q9a119 escherichia
9	1137.5	71.5	309	Q7MP36_VIBRIO	Q7mp36 vibrio vuln
10	928	58.4	309	Q6ALS2_DESPS	Q6als2 desulfohalob
11	924	58.1	320	Q4LSV5_9BURK	Q4lev5 burkholderi
12	900	56.6	308	Q4ZMV5_PSESY	Q4zmv5 pseudomonas
13	892	56.1	310	Q63M06_BURPS	Q63m06 burkholderi
14	878	55.2	312	Q8U6G0_AGRKT5	Q8u6g0 agrobacteri
15	831.5	52.3	304	Q6A8F4_PROAC	Q6a8f4 propionibac
16	773.5	48.6	307	Q88Z61_LACPL	Q88z61 lactobacill

17	581.5	36.6	308	2	Q896T9_CLOTE	Q896t9 clostridium
18	550	34.6	308	2	Q912F4_PSEAE	Q912f4 pseudomonas
19	549	34.5	323	2	Q8R1Q9_MOUSE	Q8r1q9 mus musculus
20	546	34.3	322	1	RBSK_HUMAN	Q9h477 homo sapien
21	539	33.9	311	2	Q61QD5_BRARE	Q61qd5 brachydanio
22	533	33.5	310	2	Q5M7S1_XENTR	Q5m7s1 xenopus tro
23	523	32.9	343	2	Q19133_CAEEL	Q19133 caenorhabdi
24	518	32.6	306	2	Q88316_PSESM	Q88316 pseudomonas
25	518	32.6	312	2	Q620E1_CAEER	Q620e1 caenorhabdi
26	517	32.5	306	2	Q4ZUH4_PSESY	Q4zuh4 pseudomonas
27	512.5	32.2	321	2	Q8A401_BACTN	Q8a401 bacteroides
28	507	31.9	309	2	Q982U3_RHILO	Q982u3 rhizobium 1
29	501	31.5	302	2	Q88K34_PSEPK	Q88k34 pseudomonas
30	494.5	31.1	314	2	Q7UUS5_RHOBA	Q7uus5 rhodospirell
31	493.5	31.0	306	2	Q8RD45_THETN	Q8rd45 thermoanaer
32	490.5	30.8	302	2	Q9X4M5_LACSK	Q9x4m5 lactobacill
33	488.5	30.7	309	2	Q891L9_CLOTE	Q891l9 clostridium
34	487.5	30.7	305	2	Q4KEX0_PSEF5	Q4kex0 pseudomonas
35	486.5	30.6	308	2	Q4LUI3_9BURK	Q4lui3 burkholderi
36	486.5	30.6	308	2	Q6DB84_ERWCT	Q6db84 erwinia car
37	480.5	30.2	304	2	Q6GK43_STAAR	Q6gk43 staphylococ
38	480.5	30.2	304	2	Q7A7T7_STAAN	Q7a7t7 staphylococ
39	480.5	30.2	304	2	Q99WV7_STAAM	Q99wv7 staphylococ
40	478.5	30.1	304	2	Q6GCK4_STAAS	Q6gck4 staphylococ
41	478.5	30.1	304	2	Q8NYG7_STAAN	Q8nyg7 staphylococ
42	478.5	30.1	307	2	Q5FLF7_LACAC	Q5flf7 lactobacill
43	477.5	30.0	311	2	Q4LRB7_9BURK	Q4lrb7 burkholderi
44	476.5	30.0	297	2	Q5KUX1_GEOKA	Q5kux1 geobacillus
45	476.5	30.0	304	2	Q5HJA8_STAAC	Q5hja8 staphylococ

ALIGNMENTS

RESULT 1

Q7CPF5_SALTY

ID Q7CPF5_SALTY PRELIMINARY; PRT; 306 AA.

AC Q7CPF5

DT 05-JUL-2004 (TReMBLrel. 27, Created)

DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)

DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)

DE Putative sugar kinase (SC 2.7.1.15).

GN OrderedLocustNames=STM3793;

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Salmonella.

OX NCBI_TaxID=602;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=LT2;

RX MEDLINE=21534948; PubMed=11677609; DOI=10.1038/35101614;

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,

RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,

RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,

RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,

RA Waterston R., Wilson R.K.;

RT "Complete genome sequence of Salmonella enterica serovar Typhimurium

LT2."

RL Nature 413:852-856(2001).

DR EMBL; AE008876; AAL22651.1; -; Genomic DNA.

DR GO; GO:00046301; F:kinase activity; IEA.

DR GO; GO:0004747; F:ribokinase activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR GO; GO:0006014; P:D-ribose metabolism; IEA.

DR InterPro; IPR011877; D:ribokin_bact.

DR InterPro; IPR011811; PfkB_region.

DR InterPro; IPR002139; Ribokinase.

DR Pfam; PF00294; PfkB; 1.

DR PRINTS; PR00990; RIBOKINASE.

DR TIGRFAMs; TIGR02152; D_ribokin_bact; 1.

KW Complete proteome. 306 AA; 33229 MW; 142A0289E2D1B953 CRC64;

SEQUENCE

Alignment Scores:

Pred. No.: 4.34e-129 Length: 306
Score: 1570.00 Matches: 306
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.74% Indels: 0
DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x Q7CPFS_SALTY (1-306)

Qy 1 ATGGATATCGCGGTTATTGGCTTAACATGCTGACCTTATCACCTACCAACAGATG 60
Db 1 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrTyThrAsnGlnMet 20
Qy 61 CCCAAGAGGGAACTCTGGAAGCCGCGCTTTAAATCGGCTGGGGGAAAGGG 120
Db 21 ProlysGluGlyGluThrLeuGluAlaProAlaPheIleGlyCysGlyGlyGly 40
Qy 121 GCGAACACGAGCGTGGCGCGCTAAAGCTCAATTCAAAAGTATTGATGTTCCACCAAGTG 180
Db 41 AlaAsnGlnAlaValAlaAlaAlaIleAlaIleAlaIleAlaIleAlaIleAlaIle 60
Qy 181 GCGACGATATTTTGGCCGACACACCATTCGTAATCTCGAATCTCGGGGATCAATACG 240
Db 61 GlyAspIlePheAlaAspAsnThrIleArgAsnLeuGluSerTrpGlyIleAsnThr 80
Qy 241 ACGTATGTAGAAAAGTACCTGTACACGAGCGCGTAGCGCGATTTTCGTCAACGCC 300
Db 81 ThrTyValGluIleValProCysThrSerSerGlyValAlaProIlePheValAsnAla 100
Qy 301 AACTCCAGCAACGATTCATCATCAAGCGCGCTACAAAGTTCTCTCGCCGGAAGAT 360
Db 101 AsnSerSerAsnSerIleLeuIleIleIleIleIleIleIleIleIleIleIleIle 120
Qy 361 ATCGATCGCGCGGGAAGATTTAAAAAATGCCAGCTATTGTTCTGCAATCGAAGTT 420
Db 121 IleAspArgAlaAlaGluAspLeuIleIleIleIleIleIleIleIleIleIleIle 140
Qy 421 CAGCTTGAAACGGTTTATCAGCAATAGAAATTTGCGAGAACACGGGATTTGAGTGT 480
Db 141 GlnLeuGluThrValTyHisAlaIleGluPheGlyLysIleGlyLysIleGlyVal 160
Qy 481 TTAACCTCGCGCAGCATTCAGGAATTTAGATATGCTTATGCTGTAAATCGCATTC 540
Db 161 LeuAsnProAlaProAlaLeuArgGluLeuAspMetSerTyAlaCysLysCysAspPhe 180
Qy 541 TTTGTACCTAATGAACACGAGCTGGAATATTAACCGGTATGCCAGTGATACCTATGAC 600
Db 181 PheValProAsnGluThrGluLeuGluIleLeuThrGlyMetProValAspThrTyAsp 200
Qy 601 CATATTCCGCGCAGCGGACGTTCCGTGTAGATAAAGCGCTGAACATATTATTGTACCC 660
Db 201 HisIleArgAlaAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnIleIleValThr 220
Qy 661 ATGGCGGAAGGCGCGCTGTGATGACGCTGACACAGGAAGTCCATGTTCCGCGGTTT 720
Db 221 MetGlyGluLysGlyAlaLeuTrpMetThrArgAspGlnGluValHisValProAlaPhe 240
Qy 721 AGAGTGAACGCTGTGTATACCAGCGCGCGCGATGCTTTTATCGGCTGTTTCGCGAT 780
Db 241 ArgValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheAlaHis 260
Qy 781 TACTACGTCCAGCGGGGATGTGGAAGCCGCCATGAAGAAAGCGCTCTTTCGCGCT 840
Db 261 TyrTyValGlnSerGlyAspValGluAlaAlaMetLysLysAlaValLeuPheAlaAla 280
Qy 841 TTCAGCGTCACCGGAAGGACCAATCCTTATCCAAGCATTTGACCAATTAATGAG 900
Db 281 PheSerValThrGlyLysGlyThrGlnSerSerTyProSerIleGluGlnPheAsnGlu 300
Qy 901 TATCTTTCGTTGAACGAA 918
Db 301 TyrLeuSerLeuAsnGlu 306

RESULT 2

Q8XGW5_SALTY
ID Q8XGW5_SALTY PRELIMINARY; PRT; 306 AA.
AC Q8XGW5; Q7AL29;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 10-MAY-2005 (TReMBLrel. 30, Last annotation update)
DE Putative carbohydrate kinase
GN OrderedLocusNames=STY3989, t3725;
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
CX NCBI_TaxID=601;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA DOI=10.1128/JB.185.7.2330-2337.2003;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyonanni V., Schwartz D.C., Blattner F.R.;
RA "Comparative Genomics of Salmonella enterica serovar Typhi strains Ty2
and CT18";
RL J. Bacteriol. 185:2330-2337(2003).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;
RA Parkhill J., Dougan G., James K.B., Thomson N.R., Pickard D., Wain J.,
RA Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18";
RL Nature 413:848-852(2001).
DR EMBL; AE016846; AA071218.1; -; Genomic DNA.
DR EMBL; AL627280; CAD03201.1; -; Genomic DNA.
DR HSSP; P05054; 1RKD.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0004747; F:ribokinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0006014; P:D-ribose metabolism; IEA.
DR InterPro; IPR011877; D:ribokin_bact.
DR InterPro; IPR011611; PfKB region.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; PfKB; 1.
DR PRINTS; PR00990; RIBOKINASE.
DR TIGRFAMs; TIGR02152; D:ribokin_bact; 1.
KW Complete proteome; Kinase.
SQ SEQUENCE 306 AA; 33229 MW; 142A0289E2D1B953 CRC64;

Alignment Scores:

Pred. No.: 4.34e-129 Length: 306
Score: 1570.00 Matches: 306
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.74% Indels: 0
DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x Q8XGW5_SALTY (1-306)

Qy 1 ATGGATATCGCGGTTATTGGCTCTAACATGCTGACCTTATCACCTACCAACAGATG 60
Db 1 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrTyThrAsnGlnMet 20
Qy 61 CCCAAGAGGGAACTCTGGAAGCCGCGCTTTAAATCGGCTGGGGGAAAGGG 120
Db 21 ProlysGluGlyGluThrLeuGluAlaProAlaPheIleGlyCysGlyGlyGly 40


```
Db 181 PheValProAsnGluThrGluLeuGluIleLeuThrGlyMetProValAspThrTyrAsp 200
Qy 601 CATATTCGCGAGCGGCGAGTCGCTGTAGATAAAGGCGTGAACAATATTATTGTCAAC 660
Db 201 HisIleArgAlaAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnIleIleValThr 220
Qy 661 ATGGCGGAGAAGCGCGCTGTGGATGACGGTGCACAGGAGTCCATGTTCCGCGCTTT 720
Db 221 MetGlyGluLysGlyAlaLeuTyrMetThrArgAspGlnGluValHisValProAlaPhe 240
Qy 721 AGAGTGAACGCTGTTGTATACAGCGCGCGCGCATGCTTTATCGGCTGTTTCGCGCAT 780
Db 241 ArgValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheAlaHis 260
Qy 781 TACTACGTCCAGAGCGGGGATGTGAAGCGCCCATGAAAAAGCGCTCTTTGCGCGCT 840
Db 261 TyrTyrValGlnSerGlyAspValGluAlaAlaMetLysLysAlaValLeuPheAlaAla 280
Qy 841 TTCAGCGTCACCGGAGAGGACCCCAATCCTTATCAAGCATTTGAGCAATTTAATGAG 900
Db 281 PheSerValThrGlyLysGlyThrGlnSerSerTyrProSerIleGluGlnPheAsnGlu 300
Qy 901 TATCTTTCTGTTGAACGAA 918
Db 301 TyrLeuSerLeuAsnGlu 306

RESULT 4
Q57143 SALCH PRELIMINARY; PRT; 306 AA.
ID Q57143
AC Q57143
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DE Putative sugar kinase, ribokinase family.
GN Name=bsk; OrderedLocusNames=SC3713;
OS Salmonella cholerae-suis (Salmonella enterica).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=591;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=SC-B67;
RX PubMed=15781495;
RA Chiu C.-H., Tang P., Chu C., Hu S., Bao Q., Yu J., Chou Y.-Y.,
RA Wang H.-S., Lee Y.-S.;
RT "The genome sequence of Salmonella enterica serovar Choleraesuis, a
RT highly invasive and resistant zoonotic pathogen.";
RL Nucleic Acids Res. 33:1690-1698(2005).
DR EMBL; AE017220; AAX67619.1; -; Genomic_DNA.
KW Complete proteome; Kinase.
SQ SEQUENCE 306 AA; 33187 MW; 159A0289E2D1A942 CRC64;

Alignment Scores:
Pred. No.: 1.79e-128 Length: 306
Score: 1563.00 Matches: 305
Percent Similarity: 99.67% Conservative: 0
Best Local Similarity: 99.67% Mismatches: 1
Query Match: 98.30% Indels: 0
DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x Q57143_SALCH (1-306)
Qy 1 ATGGATATCGCGGTTATGGCTTAACATGATGGACCTTATCACCTACCAACAGATG 60
Db 1 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrTyrThrAsnGlnMet 20
Qy 61 CCCAAGAGCGGGAACCTCTGAGAGCGCGCGGTTTAAATCGGCTGGCGGGAAGGG 120
Db 21 ProlysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyLysGly 40
Qy 121 GCGAACACGAGCGGTGGCGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTG 180
```

```
Db 41 AlaAsnGlnAlaValAlaAlaAlaLysLeuAsnSerLysValLeuMetLeuThrLysVal 60
Qy 181 GGCCACCATATTTTTCGCGCAACACCATTCGTAAATCTCGAATCTCGGGGATCAATACG 240
Db 61 GlyAspAspIlePheAlaAspAsnThrIleArgAsnLeuGluSerTrpGlyIleAsnThr 80
Qy 241 ACGTATGTAGAAAAAGTAGTCCGTGTACCAGCAGCGCGTAGCGCGATTTTCGTCAACGCC 300
Db 81 ThrTyrValGluLysValProCysThrSerSerGlyValAlaProIlePheValAsnAla 100
Qy 301 AACTCCAGCAACACGATTCGATCATCAAGGCGCTCAAGTTTCTCTCCCGGAGAT 360
Db 101 AsnSerSerAsnSerIleLeuIleIleLysGlyAlaAsnLysPheLeuSerProGluAsp 120
Qy 361 ATCCATCGCGCGCGGAGATTATAAAAAATGCCAGCTTATTGTTCTCAACTGGAAGTT 420
Db 121 IleAspArgAlaAlaGluAspLeuLysCysGlnLeuIleValLeuGlnLeuGluVal 140
Qy 421 CAGCTTGAACCGTATTATCAGCAATAGAAATTTGGCAAGAAACACGGGATTGAAGTTA 480
Db 141 GlnLeuGluThrValTyrHisAlaIleGluPheGlyLysLysHisGlyIleGluValLeu 160
Qy 481 TTAACCCCTCGCGCAGCATTTACGGGAATTAGATATGCTTATGCTGTAAATCGGATTTC 540
Db 161 LeuAsnProAlaProAlaLeuArgGluLeuAspMetSerTyrAlaCysLysCysAspPhe 180
Qy 541 TTTGTACTTAATGAACCGAGCTCGAAATATTAAACCGGTATGCCAGTGGATACCTATGAC 600
Db 181 PheValProAsnGluThrGluLeuGluIleLeuThrGlyMetProValAspThrTyrAsp 200
Qy 601 CATATTCGCGCAGCGCGCATTCGCTGTAGATAAAGGCGTGAACAATATTATTGTCAAC 660
Db 201 HisIleArgAlaAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnIleIleValThr 220
Qy 661 ATGGCGGAGAAGCGCGCTGTGTGATGACGGTGCACAGGAGTCCATGTTCCGCGCTTT 720
Db 221 MetGlyGluLysGlyAlaLeuTyrMetThrArgAspGlnGluValHisValProAlaPhe 240
Qy 721 AGAGTGAACGCTGTTGATACCGAGCGCGCGCATGCTTTATCGGCTGTTTCGCGCAT 780
Db 241 ArgValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheAlaHis 260
Qy 781 TACTACGTCCAGAGCGGGGATGTGAAGCGCCCATGAAAAAGCGCTCTCTTTGCCGCT 840
Db 261 TyrTyrValGlnSerGlyAspValGluAlaAlaMetLysLysAlaGlyLeuPheAlaAla 280
Qy 841 TTCAGGTCTACCGGGAAGGACCCCAATCCTCTTATCAAGCATTTGAGCAATTTAATGAG 900
Db 281 PheSerValThrGlyLysGlyThrGlnSerSerTyrProSerIleGluGlnPheAsnGlu 300
Qy 901 TATCTTTCTGTTGAACGAA 918
Db 301 TyrLeuSerLeuAsnGlu 306

RESULT 5
Q67FX7_ECOLI PRELIMINARY; PRT; 306 AA.
ID Q67FX7_ECOLI
AC Q67FX7;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Deoxyribokinase.
GN Name=dek;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=55989, AL862, and EC185;
RX PubMed=15385522; DOI=10.1128/IAI.72.10.6151-6156.2004;
RA Bernier-Februe C., Du Merle L., Turlin E., Labas V., Ordenez J.,
RA Gilles A.M., Le Bouguenec C.;
```

RT "Use of Deoxyribose by Intestinal and Extraintestinal Pathogenic
 RT Escherichia coli Strains: a Metabolic Adaptation Involved in
 RL Competitiveness";

RL Infect. Immun. 72:6151-6156 (2004).
 DR EMBL; AY298765; AAQ75098.1; -; Genomic DNA.
 DR EMBL; AY299335; AAQ83784.1; -; Genomic DNA.
 DR EMBL; AY299336; AAQ83788.1; -; Genomic DNA.
 DR GO; GO:0016301; F:kinase activity; IEA.
 DR GO; GO:0004747; F:kinase activity; IEA.
 DR GO; GO:0016740; F:ribokinase activity; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR GO; GO:0006014; P:D-ribose metabolism; IEA.
 DR InterPro; IPR011877; D:riboin bact.
 DR InterPro; IPR011611; PfKb region.
 DR InterPro; IPR002139; Ribokinase.
 DR Pfam; PF00294; PfKb; 1.
 DR PRINTS; PR00990; RIBOKINASE.
 DR TIGRFAMs; TIGR02152; D_riboin_bact; 1.
 KW Kinase.
 SQ SEQUENCE 306 AA; 33222 MW; 8A0FBFB6BFE493DB CRC64;

Alignment Scores:
 Pred. No.: 1,346-124 Length: 306
 Score: 1519.00 Matches: 294
 Percent Similarity: 98.69% Conservative: 8
 Best Local Similarity: 96.08% Mismatches: 4
 Query Match: 95.53% Indels: 0
 DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x Q67FX7_ECOLI (1-306)

QY 1 ATGGATATCGCGGTTATTGGCTTCTAACATGGTGGACCTTACCTACCAACAGATG 60
 DB 1 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrTyrThrAsnGlnMet 20
 QY 61 CCCAAGAGGGGAACTCTGAAGCGCGCGGTTTAAATCGGCTGGCGGGAAGGG 120
 DB 21 ProLysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyLysGly 40
 QY 121 GCGACACAGCGCGTGGCGCGCTAAGCTCAATTCAAAGTATTGATGTTGACCAAGTG 180
 DB 41 AlaAsnGlnAlaValAlaAlaLysLeuAsnSerLysValLeuMetLeuThrLysVal 60
 QY 181 GCGACAGATATTTTGGCGCACACACCATTCGTAATCTCGAATCTGGGGATCAATACG 240
 DB 61 GlyAspAspIlePheAlaAspAsnThrIleArgAsnLeuGluSerTyrGlyIleAsnThr 80
 QY 241 AGGTATGTAGAAAAAGTACCGTGTACACGACGCGGTAGCGCGGATTTCTGCAACGCC 300
 DB 81 ThrTyrValGluLysValProCysThrSerSerGlyValAlaProIlePheValAsnAla 100
 QY 301 RACTCCAGCAAGATTCGTGATCATCAAGGCGCTAACAGTTTCTCTCGCGCGAAGAT 360
 DB 101 AsnSerSerAsnSerIleLeuIleIleLysGlyAlaAsnLysPheLeuSerProGluAsp 120
 QY 361 ATCGATCGCGCGCGAAGATTTAAAAATGCCAGCTTATGTTCTGCAATCGGAAGTT 420
 DB 121 IleAspArgAlaAlaGluAspLeuLysLysCysLysLeuIleValLeuGlnLeuGluVal 140
 QY 421 CAGCTTGAACGGTTTATCACCAATAGAAATTGGCAAGAAACACGCGGATGAAGTGTTA 480
 DB 141 GlnLeuGluThrValTyrHisAlaIleGluPheGlyLysLysAsnGlyIleGluValLeu 160
 QY 481 TTAACCTGCGCCAGCATTACGGGATATGATATGCTTATGCTGCTGTAATGCGGATTC 540
 DB 161 LeuAsnProAlaProAlaLeuArgGluLeuAspMetSerTyrAlaCysLysCysAspPhe 180
 QY 541 TTTGTACCTAATGAACAGCTGGAATATTAAACCGGTATGCCAGTGGATACCTATGAC 600
 DB 181 PheIleProAsnGluThrGluLeuGluIleLeuThrGlyMetSerValAspThrTyrAsp 200
 QY 601 CATATTGCGGACGCGCAGTTTCGCTGGTAGATAAAGGCGCTGAACAATATTATGTCACC 660

Db 201 HisIleArgLeuAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnIleValThr 220
 QY 661 ATGGCGCAGAAAGCGCGCTGTGGATACGCGTGACCGAGGAGTCCATGTTCCGCGCTTT 720
 Db 221 MetSerGluLysGlyAlaLeuTyrMetThrArgAspGlnGluValHisValProAlaPhe 240
 QY 721 AGAGTGAACGCTGTGATACACGCGCGCGCGATGCTTTATCGGCTGTTTCGCGCAT 780
 Db 241 LysValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheSerHis 260
 QY 781 TACTACGTCCAGACGCGGATGTGAAAGCCGCCCATGAAAAAGCGCTCTCTTTGCCGCT 840
 Db 261 TyrTyrValGlnSerGlyAspValGluAlaLeuLysLysAlaAlaLeuPheAlaAla 280
 QY 841 TTCAGGTCACCGGGAAGGACCAATCTCTTATCCAGCATTGACCAATTTAATGAG 900
 Db 281 PheSerValThrGlyLysGlyThrGlnSerSerTyrProSerIleGluGlnPheAsnGlu 300
 QY 901 TATCTTTCGTTGAACGRA 918
 Db 301 PheLeuThrLeuAsnGlu 306

RESULT 6

Q8FKR4_ECOL6 PRELIMINARY; PRT; 306 AA.

AC Q8FKR4; 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Putative ribokinase.

GN OrderedLocustNames=c0331;

OS Escherichia coli O6.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=217992;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=O6:H1 / CFT073 / ATCC 700928;

RX MEDLINE=22388234; PubMed=12471157; DOI=10.1073/pnas.252597999;

RA Reick R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,

RA Wasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,

RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,

RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;

RT "Extensive mosaic structure revealed by the complete genome sequence

of uropathogenic Escherichia coli.,"

RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).

DR EMBL; AE016756; AAN78819.1; -; Genomic DNA.

DR HSP; P05054; 1RKO.

DR GO; GO:0016301; F:kinase activity; IEA.

DR GO; GO:0004747; F:ribokinase activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR GO; GO:0006014; P:D-ribose metabolism; IEA.

DR InterPro; IPR011877; D:riboin bact.

DR InterPro; IPR011611; PfKb region.

DR InterPro; IPR002139; Ribokinase.

DR Pfam; PF00294; PfKb; 1.

DR TIGRFAMs; TIGR02152; D_riboin_bact; 1.

KW Complete proteome; Kinase.

SQ SEQUENCE 306 AA; 33222 MW; 8A0FBFB6BFE493DB CRC64;

Alignment Scores:

Pred. No.: 1,346-124 Length: 306
 Score: 1519.00 Matches: 294
 Percent Similarity: 98.69% Conservative: 8
 Best Local Similarity: 96.08% Mismatches: 4
 Query Match: 95.53% Indels: 0
 DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x Q8FKR4_ECOL6 (1-306)

QY 1 ATGGATATCGCGGTTATTGGCTTCTAACATGGTGGACCTTACCTACCAACAGATG 60

Db 1 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrTyThrAsnGlnMet 20
Qy 61 CCCAAAGAGGGGAAACTCTGAGACCGCGCGGTTTAAATCGCTCGCGGCGGAAAAGGG 120
Db 21 ProLysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyGlyysGly 40
Qy 121 CGGACCAAGCGCGCGCGCGCTAAAGCTCAATCAAAAGTATTGATGTTGACCAAGTG 180
Db 41 AlaAsnGlnAlaValAlaAlaAlaLysLeuAsnSerLysValLeuMetLeuThrLysVal 60
Qy 181 GCGCAGCATATTTTTCGCGCACACACCATTCGTATCTCGAATCTCGGGGATCAATACG 240
Db 61 GlyAspAspIlePheAlaAsnThrIleArgAsnLeuGluSerTrpGlyIleAsnThr 80
Qy 241 ACGTATGTAGAAAAGTACCGGTACACAGCGCGCTAGCGCGCATTTTCTGCACGCC 300
Db 81 ThrTyValGluLysValProCysThrSerSerGlyValAlaProIlePheValAsnAla 100
Qy 301 AACTCCAGCAACAGCATTCGATCATCAAGCGCGCTACAAAGTTTCTCTCGCGGAGAT 360
Db 101 AsnSerSerAsnSerIleLeuIleLysGlyAlaAsnLysPheLeuSerProGluAsp 120
Qy 361 ATCGATCCGCGCGCGGAGAGATTAAATAAATGCCAGCTTATTGTTCTGCAACTGGAAGTT 420
Db 121 IleAspArgAlaAlaGluAspLeuLysCysLysLeuIleValLeuGlnLeuGluVal 140
Qy 421 CAGCTTGAAACCGTTTATCAGCAATAGAAATTTGGCAAGAAACACGGAATGAAGTGTTA 480
Db 141 GlnLeuGluThrValTyHisAlaIleGluPheGlyLysLysAsnGlyIleGluValLeu 160
Qy 481 TTAACCTCGCGCGCGCATACCGGAATTAGATATGCTTATGCTGTAAATGCGATTTC 540
Db 161 LeuAsnProAlaProAlaLeuArgGluLeuAspMetSerTyAlaCysLysCysAspPhe 180
Qy 541 TTTGTACCTAATGAAACCGAGCTGGAATATTAACCGGTATGCCAGTGCATACCTATGAC 600
Db 181 PheIleProAsnGluThrGluLeuGluIleLeuThrGlyMetSerValAspThrTyAsp 200
Qy 601 CATATTCGCGCAGCGCGCATGTTGCTGTAGATAAAGCGGTGAACAATATTATGTCACC 660
Db 201 HisIleArgLeuAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnIleIleValThr 220
Qy 661 ATGGCGGAGAAAGCGCGCTGTGGATGACGGTGACAGGAGTCCATGTTCCGCGTTT 720
Db 221 MetSerGluLysGlyAlaLeuTrpMetThrArgAspGlnGluValHisValProAlaPhe 240
Qy 721 AGAGTGAACGCTGTTGATACACAGCGCGCGCGCATGCTTTATCGGCTGTTTCCGCGAT 780
Db 241 LysValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheSerHis 260
Qy 781 TACTACGTCCAGAGCGGGATGTGAAGCCGCGCATGAAAAAGCCGCTCTTTCGCCGCT 840
Db 261 TyrTyValGlnSerGlyAspValGluAlaLeuLysLysAlaAlaLeuPheAlaAla 280
Qy 841 TTCAGCTCACGGGAAGGACCCCAATCCTTATCCAGCATTCGCAATTAATCAG 900
Db 281 PheSerValThrGlyLysGlyThrGlnSerSerTyProSerIleGluGlnPheAsnGlu 300
Qy 901 TATCTTTCTGTTGAACGAA 918
Db 301 PheLeuThrLeuAsnGlu 306

RESULT 7

Q70712_ECOLI PRELIMINARY; PRT; 315 AA.
AC Q70712-
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=536;
RX PubMed=15385503; DOI=10.1128/IAI.72.10.5993-6001.2004;
RA Schneider G., Dobrindt U., Bruggemann H., Nagy G., Janke B.,
Blum-Oehler G., Buchrieser C., Gottschalk G., Emody L., Hacker J.;
"The Pathogenicity Island-Associated K15 Capsule Determinant Exhibits
a Novel Genetic Structure and Correlates with Virulence in
RT Uropathogenic Escherichia coli Strain 536.";
RL Infect. Immun. 72:5993-6001(2004).
DR EMBL; AJ617685; CAB85173.1; -; Genomic DNA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0004747; F:ribokinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0006014; P:D-ribose metabolism; IEA.
DR InterPro; IPR011877; D_ribokin_bact.
DR InterPro; IPR011611; PfKB region.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; Ffkb; 1.
DR PRINTS; PR00990; RIBOKINASE.
DR TIGRFAMs; TIGR02152; D_ribokin_bact; 1.
KW Hypothetical protein.
SQ SEQUENCE 315 AA; 34327 MW; 8A9721EB93A15D59 CRC64;

Alignment Scores:

Pred. No.: 1.34e-124 Length: 315
Score: 1519.00 Matches: 294
Percent Similarity: 98.69% Conservative: 8
Best Local Similarity: 96.08% Mismatches: 4
Query Match: 95.53% Indels: 0
DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x Q70712_ECOLI (1-315)

Qy 1 ATGGATATCGCGGTATTGGCTTAACATGCTGGACCTTATCACCTACCAACCCAGATG 60
Db 10 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrTyThrAsnGlnMet 29
Qy 61 CCCAAAGAGGGGAAACTCTGAGACCGCGCGTAAATCGCTCGCGGCGGAAAAGGG 120
Db 30 ProLysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyGlyysGly 49
Qy 121 CGGACCAAGCGCGCGCGCTAAAGCTCAATCAAAAGTATTGATGTTGACCAAGTG 180
Db 50 AlaAsnGlnAlaValAlaAlaLysLeuAsnSerLysValLeuMetLeuThrLysVal 69
Qy 181 GCGCAGCATATTTTTCGCGCACACACCATTCGTATCTCGAATCTCGGGGATCAATACG 240
Db 70 GlyAspAspIlePheAlaAsnThrIleArgAsnLeuGluSerTrpGlyIleAsnThr 89
Qy 241 ACGTATGTAGAAAAGTACCGGTACACAGCGCGCTAGCGCGCATTTTCTGCACGCC 300
Db 90 ThrTyValGluLysValProCysThrSerSerGlyValAlaProIlePheValAsnAla 109
Qy 301 AACTCCAGCAACAGCATTCGATCATCAAGCGCGCTACAAAGTTTCTCTCGCGGAGAT 360
Db 110 AsnSerSerAsnSerIleLeuIleLysGlyAlaAsnLysPheLeuSerProGluAsp 129
Qy 361 ATCGATCCGCGCGGAGAGATTAAATAAATGCCAGCTTATTGTTCTGCAACTGGAAGTT 420
Db 130 IleAspArgAlaAlaGluAspLeuLysCysLysLeuIleValLeuGlnLeuGluVal 149
Qy 421 CAGCTTGAAACCGTTTATCAGCAATAGAAATTTGGCAAGAAACACGGAATGAAGTGTTA 480
Db 150 GlnLeuGluThrValTyHisAlaIleGluPheGlyLysLysAsnGlyIleGluValLeu 169
Qy 481 TTAACCTCGCGCGCGCATACCGGAATTAGATATGCTTATGCTGTAAATGCGATTTC 540
Db 170 LeuAsnProAlaProAlaLeuArgGluLeuAspMetSerTyAlaCysLysCysAspPhe 189


```

Qy 541 TTTGTACTAATGAAACCGAGCTGGAATATTAACCGGTATGCCAGTGTGATACCTATGAC 600
Db 190 PheileProAenGluThrGluLeuGluileLeuThrGlyMetSerValAspThrTyrAsp 209
Qy 601 CATATTCGGCAGCGGCGAGTTCGCTGCTAGTAAAGCGGTGAACAATATTATTGTCCACC 660
Db 210 HisileArgLeuAlaAlaArgSerLeuValAspLysGlyLeuAenAsnilelleValThr 229
Qy 661 ATGGCGGAGAAAGCGCGCTGTGGATGACGGTGTGACACGAGGAGTCCATGTTCCGCGCTTT 720
Db 230 MetSerGluLysGlyAlaLeuTyrMetThrArgAspGlnGluValHisValProAlaPhe 249
Qy 721 AGAGTGAACGCTGTGTGATACACAGCGCGCGCGGATGCGCTTATCGGCTGTTCCGCGAT 780
Db 250 LysValAenAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheSerHis 269
Qy 781 TACTACGTCCAGAGCGGGGATGTGAAGCGCGCCATGAAAGGCGTCTCTTTCGCGCT 840
Db 270 TyrTyrValGlnSerGlyAspValGluAlaLeuLysLysAlaLeuPheAlaAla 289
Qy 841 TTCAGCGTCACCGGAAAGGACCAATCCCTCTTATCCAAAGCATTTGAGCAATTTAATGAG 900
Db 290 PheSerValThrGlyLysGlyThrGlnSerSerTyrProSerIleGluGlnPheAsnGlu 309
Qy 901 TATCTTTCCGTTGAACGAA 918
Db 310 PheLeuThrLeuAsnGlu 315

RESULT 8
Q9A119_ECOLI PRELIMINARY; PRT; 314 AA.
AC Q9A119;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein ORF4.
GN Name=ORF4;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_taxid=562;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AL862;
RX MEDLINE=21101044; PubMed=11159989; DOI=10.1128/IAI.69.2.937-948.2001;
RA Lalioui L., Le Boughenec C.C.;
RT "afa-8 Gene cluster is carried by a pathogenicity island inserted into
RT the tRNA (Phe) of human and bovine pathogenic Escherichia coli
RT isolates.";
RL Infect. Immun. 69:937-948(2001).
DR EMBL; AF286671; AAK27331.1; -; Genomic_DNA.
DR HSSP; P05054; 1GQT.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0004747; F:ribokinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0006014; P:D-ribose metabolism; IEA.
DR InterPro; IPR011877; D_ribokin_bact.
DR InterPro; IPR011611; PfkB_region.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; PfkB; 1.
DR PRINTS; PR00990; RIBOKINASE.
DR TIGRFam; TIGR02152; D_ribokin_bact; 1.
KW Hypothetical protein.
SQ SEQUENCE 314 AA; 34415 MW; 2D5F7342C817E3BB CRC64;

```

Alignment Scores:

```

Pred. No.: 2,99e-117 Length: 314
Score: 1435.50 Matches: 287
Percent Similarity: 95.16% Conservatives: 8
Best Local Similarity: 92.58% Mismatches: 7
Query Match: 90.28% Indels: 9
DB: 2 Gaps: 2

```

```

US-10-049-750-11 (1-921) x Q9A119_ECOLI (1-314)
Qy 1 ATGGATATCGCGGTATTGGCTCTTAACATGCTGAGCACTTATCACCTACCAACCAAGATG 60
Db 10 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrTyrThrAsnGlnMet 29
Qy 61 CCCAAAGAGGGGAAACTCTGGAAGCCGCGGCTTAAATCGGCTCGCGGGAAGGG 120
Db 30 ProLysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyGly 49
Qy 121 GCGAACACGCGCGTGGCGCGCTAAGCTCAATTCAAAGTATTGATCTTCACCAAGATG 180
Db 50 AlaAsnGlnAlaValAlaAlaLysLeuAsnSerLysValLeuMetLeuThrLysVal 69
Qy 181 GCGACGATATTTTGGCGCAACACCATTCGTAAATCTCGAATCTCGGGGATCAATACG 240
Db 70 GlyAspAspIlePheAlaAspAsnThrIleArgAsnLeuGluSerTyrGlyIleAsnThr 89
Qy 241 ACGTATGTAGAAAAGTACCGTGTACAGCAGCGCGTAGCGCGGATTTCTGTCACGCC 300
Db 90 ThrTyrValGluLysValProCysThrSerSerGlyValAlaProIlePheValAsnAla 109
Qy 301 AACTCCAGCAACAGCATTTCTCATCAAGCGCGCTAACAAAGTTTCTCTCGCGGAAGAT 360
Db 110 AsnSerSerAsnSerIleLeuIleLysGlyAlaAsnLysPheLeuSerProGluAsp 129
Qy 361 ATCGATCGCGCGGGAAGATTAAATAAAATGCCAGCTTATTGTTTCTCAACTGGAAGTT 420
Db 130 IleAspArgAlaAlaGluAspLeuLysLysCysLysLeuIleValLeuGlnLeuGluVal 149
Qy 421 CAGCTTGAACCGTTTATCAGCAATAGAAATTTGGCAAGAAACACGGGATGAACTGTTA 480
Db 150 GlnLeuGluThrValTyrHisAlaIleGluPheGlyLysLysAsnGlyIleGluValLeu 169
Qy 481 TTAAA-----CCCTCGCCAGCATTTACGGGAATTAGATATGTCTTATGCTCTGATAA 530
Db 170 LeuIleArgLeuArgHisCysValSer-----ArgTyrVal-TyrAlaCysLys 185
Qy 531 ATGCGATTCTTTGTACCTAATGAACCGAGCTGGAATATTAAACCGGTATGCCAGTGA 590
Db 185 sCysAspPhePheIleProAsnGluThrGluLeuGluIleLeuThrGlyMetSerValAs 205
Qy 591 TACCTATGACCATATTCCGCGAGCGGACGTCGTGCTGTAGATAAAGGGCTGAACAATAT 650
Db 205 pThrTyrAspHisileArgLeuAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnI 225
Qy 651 TATTGTCCACCATGGCGAGAAAGCGCGCTGTGGATGACGCGTGACAGGAGTCCATGCT 710
Db 225 eileValThrMetSerGluLysGlyAlaLeuThrMetThrArgAspGlnGluValHisVa 245
Qy 711 TCCGCGGTTTAGAGTGAACGCTGTTGTATACACGCGCGCGCGCATGCTTTATCGGCTG 770
Db 245 lProAlaPheLysValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCy 265
Qy 771 TTTCCGCGCATTTACTAGTCCAGAGCGGGGATGTGGAACCGCCATGAAAAGCCGCTCT 830
Db 265 sPheSerHisTyrTyrValGlnSerGlyAspValGluAlaLeuLysLysLysAlaAlaLe 285
Qy 831 CTTTCCGCGCTTTCAGCGTCACCGGAAAGGACCAATCTCTTATCAAGCATTTGAGCA 890
Db 285 uPheAlaAlaPheSerValThrGlyLysGlyThrGlnSerSerTyrProSerIleGluGl 305
Qy 891 ATTTAATGAGTATCTTCTGTTGAACGAA 918
Db 305 nPheAsnGluPheLeuThrLeuAsnGlu 314

RESULT 9
Q7MP36_VIBVY PRELIMINARY; PRT; 309 AA.
ID Q7MP36_VIBVY
AC Q7MP36_VIBVY
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)

```



```

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Sugar kinase.
GN OrderedLocusNames=V0528;
OS Vibrio vulnificus (strain VJ016).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=196600;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=1465695; DOI=10.1101/gr.1295503;
RA Chen C.-Y., Wu K.-M., Chang Y.-C., Chang C.-H., Tsai H.-C.,
RA Liao T.-L., Liu Y.-M., Chen H.-J., Shen A.B.-T., Li J.-C., Su T.-L.,
RA Shao C.-P., Lee C.-T., Hor L.-I., Tsai S.-F.;
RT "Comparative genome analysis of Vibrio vulnificus, a marine
RT pathogen.";
RL Genome Res. 13:2577-2587(2003).
DR EMBL; BA000037; BAC93292.1; -; Genomic_DNA.
DR HSSP; P05054; 1GQT.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0004747; F:ribokinase activity; IEA.
DR GO; GO:0006014; P:D-ribose metabolism; IEA.
DR InterPro; IPR011877; D_ribokin_bact.
DR InterPro; IPR011611; PfKB_region.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; PfKB; 1.
DR PRINTS; PR00990; RIBOKINASE.
DR TIGRFAMs; TIGR02152; D_ribokin_bact; 1.
DR Complete proteome; Kinase.
KW SEQUENCE 309 AA; 33507 MW; EBB8ACF60CB6C44D CRC64;
SQ
Alignment Scores:
Pred. No.: 4, 98e-91 Length: 309
Score: 1137.50 Matches: 218
Percent Similarity: 86.6% Conservative: 42
Best Local Similarity: 72.67% Mismatches: 39
Query Match: 71.54% Indels: 1
Gaps: 2
DB: 1

US-10-049-750-11 (1-921) x Q7MP36_VIBVY (1-309)
QY 7 ATCCGGTATTGGCTTAACATGGTGGACCTTATCACCTACCAACAGATGCCAAA 66
Db 4 IleAlaValIleGlySerAsnMetValAspLeuIleSerTyThrAspArgMetProlys 23
QY 67 GAAGGGAACTCGAAGACCGCGCTTTAAATCGCTCGCGGGAAGGGCGAAC 126
Db 24 GluGlyGluThrLeuGluAlaProSerPheLysMetGlyCysGlyGlyLeuAlaAsn 43
QY 127 CAGGCGTGGCGCGCTTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGAC 186
Db 44 GlnAlaValAlaAlaAlaLysMetGlyAlaAspValValMetValSerLysValGlyAsp 63
QY 187 GATATTTTGGCACAACACCATTCGTATCTCGAATCTCGGGGATCAATACGACGAT 246
Db 64 AspMetPheAlaAspAsnThrIleArgAsnPheGlnSerTyThrGlyLeuAsnThrGlnTy 83
QY 247 GTAGAAAAGTACCTGTACAGGACGCGCTAGCGCGATTTGTCGTCACGCCCACTCC 306
Db 84 ValSerLysValProGlnThrSerSerGlyValAlaProIlePheValAsnSerThrSer 103
QY 307 AGCACACATTCATCATCAAGCGCTCAACAGTTTCTCTCCCGGAAGATATCGAT 366
Db 104 GlnAsnSerIleLeuIleLysGlyAlaAsnGluPheLeuLysProAspAspLeu 123
QY 367 CGCGCGCGGAGAGATTAAAAAATGCCAGCTTATTGTTCTGCAACTCGAAGTTCAGCTT 426
Db 124 LysAlaGluSerThrLeuValGluCysSerLeuIleValLeuGlnGluValProLeu 143
QY 427 GAACCGTTTATCAGCAATAGAAATTTGGCAAGAACCGGATTCAGTCTTATTAAC 486
Db 144 GluThrValTyAlaAlaIleGluPheGlyAsnLysHisSerIleProValLeuLeuAsn 163
QY 487 CTGCGCCGACGATTACGGGAATTAGATATGTTATGCTGTAAATGCGATTTCTTTGTA 546

```

```

Db 164 ProAlaProAlaValProGluLeuAspIleGluTyAlaCysArgCysAspPheVal 183
QY 547 CCTAATGAAACCGAGCTGGAAATATTAAACCGGTATGCCAGTGGATACCTATGACCATATT 606
Db 184 ProAsnGluThrGluLeuGluIleLeuValAsnLysProValGluThrValGluGlnIle 203
QY 607 CGCGCAGCGGCACGTTTCGCTGGTAGATAAAGGCTGAACCAATATTATTGTCCACATGGC 666
Db 204 LysGluAlaAlaThrIleLeuLeuAsnLysGlyLeuAsnAlaIleValThrMetGly 223
QY 667 GAGAAAGCGCGCTGTGGATGACCGCGTAC---CAGGAAGTCCATGTTCGGCGTTTAGA 723
Db 224 SerLysGlyAlaLeuTrpLeuSerLysAspGlyLysAspValPheIleGluProThrLys 243
QY 724 GTGAACGCTGTGATACACGCGCGCGCGGCGATGCGCTTTATCGGCTGTTTCGCGCATTAC 783
Db 244 ValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheSerHisTy 263
QY 784 TACGTCCAGACGCGGATGTGGAAGCGCCCATGAAAAAGCCGTCCTCTTTGCCGCTTTC 843
Db 264 PheMetGlnThrGlyAspValGlnLysSerLeuGluLysAlaSerLeuPheAlaAlaPhe 283
QY 844 AGCTCACCGGGAAGCACCACCAATCTCTTATCCAGCATTTAGCAATTTAATGAGTAT 903
Db 284 SerValThrGluLysGlyThrGlnPheSerTyThrGlnPheGluGlnPheGluGluPhe 303

RESULT 10
Q6ALS2_DESPS
ID Q6ALS2_DESPS PRELIMINARY; PRT; 309 AA.
AC Q6ALS2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Related to ribokinase.
GN OrderedLocusNames=DPI974;
OS Desulfotalea psychrophila.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfobacteriales;
OC Desulfobulbaceae; Desulfotalea.
OX NCBI_TaxID=84980;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=LSV54 / DSM 12343;
RX PubMed=15305914; DOI=10.1111/j.1462-2920.2004.00665.x;
RA Rabus R., Ruepp A., Frickey T., Rattei T., Fartmann B., Stark M.,
RA Bauer M., Zibat A., Lombardot T., Becker I., Amann J., Gallner K.,
RA Teeling H., Leuschner W.D., Gloeckner F.-O., Lupas A.N., Amann R.,
RA Klenk H.-P.;
RT "The genome of Desulfotalea psychrophila, a sulfate-reducing bacterium
RT from permanently cold Arctic sediments.";
RL Environ. Microbiol. 6:887-902(2004).
DR EMBL; CR522870; CAG36703.1; -; Genomic_DNA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0004747; F:ribokinase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0006014; P:D-ribose metabolism; IEA.
DR InterPro; IPR011877; D_ribokin_bact.
DR InterPro; IPR011611; PfKB_region.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; PfKB; 1.
DR TIGRFAMs; TIGR02152; D_ribokin_bact; 1.
KW Complete proteome; Kinase.
SQ SEQUENCE 309 AA; 32997 MW; 4A16AD3CC4C86267 CRC64;
Alignment Scores:
Pred. No.: 1, 36e-72 Length: 309
Score: 928.00 Matches: 183
Percent Similarity: 74.92% Conservative: 41
Best Local Similarity: 61.20% Mismatches: 75
Query Match: 51.36% Indels: 0
Gaps: 2
DB: 0

```

US-10-049-750-11 (1-921) x Q6ALS2_DESPS (1-309)

```
QY 7 ATCGCGGTTATTGGCTTAACATGTGGACCTTATCACCACCAACAGATGCCAAA 66
DB 7 IleaValIleGlySerAsnMetValAspLeuValThrTyThrAspArgMetProVal 26
QY 67 GAAGGGGAAACTCTGAAGCGCGCGGTTTAAATCGGCTGGCGGAAAGGGCGAAC 126
DB 27 AlaGlyGluThrLeuGluAlaProAspPheAspLeuGlyPheGlyLysGlyAlaAsn 46
QY 127 CAGCGCGTGGCGCGCTTACGATCAATCAAAAGTATTGATCTGACCAAGTGGCGAC 186
DB 47 GlnAlaIleAlaAlaLysLeuGlyGlyGluValLeuMetLeuSerLysValGlyAsp 66
QY 187 GATATTTTTCGCGACACACCATTCGTATCTCGATCTCGATCTCGATCAATCAACGAT 246
DB 67 AspValPheGlyProAsnThrArgAlaAsnPheValAlaAsnGlyIleAspAlaArgTy 86
QY 247 GTAGAAAAGTAGTCGTGTACACGACGCGCGTGGACGATTTTTCGTCAACGCCAATCC 306
DB 87 ValGluThrAlaAlaGlyLeuSerGlyValAlaProIlePheValAspAlaLysGly 106
QY 307 AGCAACAGCATCTGATCATCAAGCGCTAACAAGTTCTCTCGCGGGAAGATATCGAT 366
DB 107 GlnAsnSerIleLeuIleValLysGlyAlaAsnLysHisLeuSerProValAspValAsp 126
QY 367 CGCGCGCGGAGAGATTAAATAAATCGCAGCTTATTGTTCTCAACTGGAAGTTCAGCT 426
DB 127 ArgAlaIleAspIleArgAlaCysAspLeuIleValMetGlnLeuGlySerLeu 146
QY 427 GAAACGGTTTATCAGCAATAGATATTTGGCAAGAAACACGCGGATTCAGTGTATTAAAC 486
DB 147 GluThrValTyThrValIleGluLeuGlyLeuArgGluHisIleProValLeuLeuAsn 166
QY 487 CTGCGCCAGCATATCGGGAATTAGATATGCTTATGCTGCTGTAATGCAATTCGATTTCT 546
DB 167 ProAlaProAlaValAlaGlyLeuAspMetGluLysIleCysGlnLeuAspMetLeuVal 186
QY 547 CCTAATGAACCGAGCTGGAATATTTAACCGGTATGCCAGTGGATACCTATCACCATAT 606
DB 187 ProAsnGluThrGluLeuGluIleuThrGlyMetProValGlnThrLeuGluGlnIle 206
QY 607 CGCGACGCGCGCTGCTGCTGTAGATAAAGGCGTGAACAATATTATTGTCACCATGGGC 666
DB 207 GlnAlaAlaAlaLysPheLeuIleAspAlaGlyIleLysLysValIleValThrMetGly 226
QY 667 GAGAAAGCGCGCTGTGGATGACGCGTGACAGGAAGTCCATGTTCCGCGGTTTAGATG 726
DB 227 SerLysGlyAlaLeuLeuValThrGluLysGluMetThrSerValProCysProGlnVal 246
QY 727 AACGCTGTTGATACACGCGCGCGGCGATGCTTATCGGCTGTTTCGCGCATCTACTAC 786
DB 247 LysAlaLysAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheAlaLysHisTy 266
QY 787 GTCCAGAGCGGAGTGTGAAGCGCGCATGAAAGCGCTCTCTTTCCGCTTTTCAGC 846
DB 267 ValGluGlyGlyGluLeuLeuProAlaMetGluGluAlaValIleTyAlaSerLeuSer 286
QY 847 GTCACGGGAAGGCAACCAATCTCTTATCCAAGCAATTCAGCAATTAATGAGTAT 903
DB 287 ThrThrArgProGlyThrGlnLysSerTyAlaAspIleAspGlnPheGluAlaTy 305
```

RESULT 11

```
Q4LSV5_9BURK
ID Q4LSV5_9BURK PRELIMINARY; PRT; 320 AA.
AC Q4LSV5;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Carbohydrate kinase, PfKB.
GN ORFNames=Bcen2424DRAFT_3825;
OS Burkholderia cenocepacia HI2424.
```

```
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; Burkholderia cenocepacia complex.
OX NCBI_TaxID=331272;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HI2424;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Istrati S., Pittluck S., Richardson P.;
RT "Sequencing of the draft genome assembly of Burkholderia cenocepacia
RT HI2424.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HI2424;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Burkholderia cenocepacia
RT HI2424.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AALH01000019; EAM19121.1; -; Genomic_DNA.
KW Kinase.
SQ SEQUENCE 320 AA; 33910 MW; 9A1441345CD3FA07 CRC64;
```

Alignment Scores:

Align. No.:	3,078-72	Length:	320
Score:	924.00	Matches:	183
Percent Similarity:	75.33%	Conservative:	43
Best Local Similarity:	61.00%	Mismatches:	74
Query Match:	58.11%	Indels:	0
DB:	2	Gaps:	0

US-10-049-750-11 (1-921) x Q4LSV5_9BURK (1-320)

```
QY 7 ATCGCGGTTATTGGCTTAACATGTGGACCTTATCACCACCAACAGATGCCAAA 66
DB 16 IleaValIleGlySerAsnMetValAspLeuValThrTyThrAspArgMetProAla 35
QY 67 GAAGGGGAAACTCTGAAGCGCGCGGTTTAAATCGGCTGGCGGAAAGGGCGAAC 126
DB 36 ArgGlyGluThrLeuGluAlaProAsnPheGluLeuGlyCysGlyLysGlyAlaAsn 55
QY 127 CAGCGCGTGGCGCGCTGAAGCTCAATCAAAAGTATTGATCTGACCAAGTGGCGAC 186
DB 56 GlnAlaValAlaAlaAlaArgLeuGlyAlaArgValValMetValThrLysValGlyAsp 75
QY 187 GATATTTTTCGCGACACACCATTCGTATCTCGAATCTCTGGAATCTCTGGGGATCAATACGAGTAT 246
DB 76 AspValPheAlaAspAsnThrIleArgAsnPheGluArgGluGlyIleAspThrThrHis 95
QY 247 GTAGAAAAGTAGTCGTGTACACGACGCGCGTAGCGCGATTTTCGTCAACGCCAATCC 306
DB 96 ValArgLysValAlaGlyValProSerGlyValAlaProIlePheValGluProAspSer 115
QY 307 AGCAACAGCATCTGATCATCAAGCGCGCTAAAGTTCCTCTCGCGGGAAGATATCGAT 366
DB 116 SerAsnSerIleLeuIleValLysGlyAlaAsnArgHisLeuGlnProAlaAspIleAsp 135
QY 367 CGCGCGCGGAGAGATTTAAATAAATGCCAGCTTATGTTCTGCAACTGGAAGTTCAGCTT 426
DB 136 AlaAlaAlaProMetLeuAlaGluCysAlaLeuIleValLeuGlnLeuGluLeu 155
QY 427 GAAACGGTTTATCAGCAATAGATTTGGCAAGAAACACGCGGATTCAGTGTATTAAAC 486
DB 156 AspThrValTyThrHisAlaIleGluPheGlyAlaArgHisGlyIleProValLeuLeuAsn 175
QY 487 CTGCGCGCAGCATTACGGGAATTAGATATGCTTATGCTGCTGTAATGCGATTTCTTTGTA 546
DB 176 ProAlaProAlaValAlaAspLeuAspPheGluArgGlyIleArgSerValGluPheVal 195
```

```
Qy 547 CCTAATGAACCGAGCTGAAATATTAAACCGGTATGCCAGTGGATACCTATGACCATATT 606
Db 196 ProAsnGluThrGluLeuAlaIleValSerGlyLeuProValAspSerArgGluSerAla 215
Qy 607 CGCGCAGCGCGCTGCTGGTGTAGATAAAGGCTGAAACAATATTATTGTCCACATGGGC 666
Db 216 ThrArgAlaAlaGluAlaLeuValAlaArgGlyLeuLysHisValLeuValThrLeuGly 235
Qy 667 GAGAAAGCGCGCTGTGATGACGCGTACCGAGGAGTCCATGTTCCGCGCTTTAGAGTG 726
Db 236 SerAsnGlySerLeuLeuValSerArgAspGlyValHisHisValProGlyValProVal 255
Qy 727 AACGCTGTGTATACCGAGCGCGCGCATGCTTTATCGGCTGTTTCGCGCATTTACTAC 786
Db 256 AspAlaArgAspThrThrGlyAlaGlyAspAlaTyrlleGlyCysPheAlaArgCysTyr 275
Qy 787 GTCCAGAGCGGGAGTGGAGCGCCCATGAAAAAGCGCTCTCTTTGCGCTTTTCAGC 846
Db 276 AlaAlaSerArgAspAlaValAspAlaMetArgTyrAlaSerAlaTyrAlaAlaHisSer 295
Qy 847 GTCACCGGGAAGGACCCCAATCTCTTATCCAGCATTTGAGCAATTTAATGAGTATCTT 906
Db 296 ValThrGlyLeuGlyThrGlnLysSerTyrAlaAspAlaAlaThrPheGluArgPheLeu 315

RESULT 12
Q4ZMV5_PSESYPRELIMINARY; PRT; 308 AA.
ID Q4ZMV5_PSESYPRELIMINARY; PRT; 310 AA.
AC Q4ZMV5;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Carbohydrate kinase, PfkB.
GN ORFNames=FeyT_4487;
OS Pseudomonas syringae pv. syringae B728a.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=205918;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B728a;
RG DOE Joint Genome Institute;
RA Chain P., Larimer F., DiBartolo G., Copeland A., Lykidis A., Trong S.,
RA Nolan M., Goldsman E., Thiel J., Malfatti S., Lapidus A., Detter J.C.,
RA Land M., Richardson P.M., Kyrpides N.C., Ivanova N.;
RT "Comparison of two complete genome sequences of Pseudomonas syringae
RT pv. syringae B728a and pv. tomato DC3000."
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2005).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B728a;
RA Loper J.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B728a;
RA Feil H., Feil W.S., Lindow S.E.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; CP000075; AY39517.1; -; Genomic DNA.
DR GO; GO:0016301; F:kinase activity; IEA.
KW Kinase.
SQ SEQUENCE 308 AA; 32410 MW; A3FFD98D090D2432 CRC64;

Alignment Scores:
Pred. No.: 3.94e-70 Length: 308
Score: 900.00 Matches: 179
Percent Similarity: 73.00% Conservative: 40
Best Local Similarity: 59.67% Mismatches: 81
Query Match: 56.60% Indels: 0
Dbs: 2 Gaps: 0

US-10-049-750-11 (1-921) x Q4ZMV5_PSESYP (1-308)
Qy 7 ATCGCGGTTATTGGCTCTAACAATGTTGGAGCTTATCACCTACCAACGAGATGCCCAAA 66
```

```
Db 4 IleAlaValIleGlySerAsnMetValAspLeuIleThrTyrlleGluArgMetProAla 23
Qy 67 GAAGGGGAACACTCGGAAGCGCGCTTTAAATCGCTGCGCGGAAAAAGGCGCGAAC 126
Db 24 GlnGlyLthrLeuGluAlaProArgPheAlaMetGlyCysGlyGlyValAlaAsn 43
Qy 127 CAGCGCTGGCGCGCTCAAGCTCAATTCAAAAGTATTGTATGTTGACCAAGTGGCGAC 186
Db 44 GlnAlaAlaAlaAlaLeuLeuGlyAlaAspValLeuMetLeuSerLysValGlyAsp 63
Qy 187 GATATTTTGGCGCAACACCATTCGTAATCTCGAATCTCGGGGATCAATACGACGTAT 246
Db 64 AspAlaPheAlaAspThrThrLeuAlaAsnPheGlnArgLeuGlyIleAspThrArgPhe 83
Qy 247 GTAGAAAAGTACCGTGTACCAGCAGCGCGCTAGCGCGATTTCCTCGTCAACGCCAATCC 306
Db 84 ValGluArgValProGlyValSerSerGlyValAlaProIlePheValGlnGluAsnSer 103
Qy 307 AGCAACAGCATTCGTGATCATCAAAAGCGCTAAACAAGTTTCTCTCCGCGAAGATATCGAT 366
Db 104 HisAsnSerIleLeuIleValLysGlyAlaAsnAlaHisLeuSerProAlaAspIleAsp 123
Qy 367 CGCGCGCGGGAAGATTAAAAAAATGCCAGCTTATTGTTCTGCAACTCGGAAGTTCAGCTT 426
Db 124 ArgAlaGluAlaGlnLeuArgAspCysAlaLeuIleValLeuGlnLeuGluValGluLeu 143
Qy 427 GAAACGGTTTATCACGCAATAGAATTGGCAAGAAACACCGGATTGAAGTGTATTAAAC 486
Db 144 AlaThrValTyrHisAlaIleAlaPheAlaHisArgHisAlaIleProValLeuLeuAsn 163
Qy 487 CTGCGCGCAGCATTCACGGAATTAGATATGCTTATGCTGTAAATGCGATTTCTTTGTA 546
Db 164 ProAlaProAlaLeuAlaGlyLeuSerArgGluHisLeuAlaGlnLeuAspPheLeuIle 183
Qy 547 CCTAATGAACCGAGCTGGAATATTAAACCGGTATGCCAGTGGATACCTATGACCATATT 606
Db 184 ProAsnGluSerGluLeuAlaLeuIleSerGlyGlnProValThrCysAlaGluGluAla 203
Qy 607 CGCGCAGCGCGCTGCTGCTAGATAAAGCGCTGAACAATATTATTGTCCACATGGGC 666
Db 204 AlaGlnAlaAlaArgGlnLeuValAlaLeuGlyValArgHisValIleValThrLeuGly 223
Qy 667 GAGAAAGCGCGCTGTGGATGACGCGTACAGGAGGAGTCCATGTTCCGCGCTTTAGAGTG 726
Db 224 GluGlnGlyAlaLeuTyrValGlyAlaGluGlyGluTrpGlnValProGlyValLysVal 243
Qy 727 AACGCTGTGTATACCGCGCGCGCGATCCCTTTATCGGCTGTTTCGCGCATTTACTAC 786
Db 244 GlnAlaArgAspThrThrGlyAlaGlyAspAlaPheIleGlyCysPheAlaArgHisTrp 263
Qy 787 GTCCAGAGCGGGGATGGGAAGCGCCCATGAAAAAGCGCTCTCTTTGCGCGCTTTACAGC 846
Db 264 SerGlnAspGlyAsnIleArgGlnAlaMetHisGlnAlaValAlaTyrSerAlaCysSer 283
Qy 847 GTCACCGGGAAGGACCCCAATCTCTTATCCAGCATTTGAGCAATTTAATGAGTATCTT 906
Db 284 ValThrGlyLeuGlyThrGlnSerSerTyrProAspAlaSerThrPheAlaArgPheLeu 303

RESULT 13
Q63M06_BURPSPRELIMINARY; PRT; 310 AA.
ID Q63M06_BURPSPRELIMINARY; PRT; 310 AA.
AC Q63M06;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Putative ribokinase.
GN OrderedLocusNames=BPS0853;
OS Burkholderia pseudomallei (pseudomonas pseudomallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; pseudomallei group.
OX NCBI_TaxID=28450;
RN [1]
```

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K96243;
RX PubMed=15377794; DOI=10.1073/pnas.04030210101;
RA Holden M.T.G., Titball R.W., Peacock S.J., Cerdeno-Tatrago A.-M.,
RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.L.,
RA Bentley S.D., Sebahia M., Thomson N.R., Bason N., Beacham I.R.,
RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,
RA Chillingworth T., Cronin A., Crosssett B., Davis P., DeShazer D.,
RA Feltwell T., Fraser A., Hance Z., Hauser H., Holroyd S., Jagels K.,
RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,
RA Rabinowitsch E., Rutherford K., Sanders M., Simmonds M.,
RA Songsivilai S., Stevens K., Tumapa S., Vesaratchavest M.,
RA Whitehead S., Yeats C., Barrell B.G., Oyston P.C.F., Parkhill J.;
RT "Genomic plasticity of the causative agent of melioidosis,
RT Burkholderia pseudomallei";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245 (2004).
DR EMBL; BX571966; CAH38316.1; -; Genomic_DNA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0004747; F:ribonuclease activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0006014; P:D-ribose metabolism; IEA.
DR InterPro; IPR011877; D:ribokin_bact.
DR InterPro; IPR011611; PfkB_region.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; PfkB; 1.
DR PRINTS; PR00990; RIBOKINASE.
DR TIGRFAMs; TIGR02152; D_ribokin_bact; 1.
DR Complete proteome; Kinase.
KW SEQUENCE 310 AA; 32606 MW; A58CAE50FC6754C CRC64;
Alignment Scores:
Pred. No.: 2e-69 Length: 310
Score: 892.00 Matches: 170
Percent Similarity: 76.00% Conservative: 58
Best Local Similarity: 55.67% Mismatches: 72
Query Match: 56.10% Indels: 0
Gaps: 0
DB: 2
US-10-049-750-11 (1-921) x Q63M06_BURPS (1-310)
QY 7 ATCGCGGTATTGGCTTAACATGTGGAGCTTATCACTACCAACAGATGCCAAA 66
Db 4 TleAlaValIleGlySerAsnMetValAspLeuValThrTyValThrArgMetProAla 23
QY 67 GAAGGGAAACTCTGGAACCGCGCGGTAAATCGGTGGCGGGAAGGGCGAC 126
Db 24 AspGlyGluThrLeuGluAlaProAsnPheGluLeuGlyCysGlyGlyAlaAsn 43
QY 127 CAGGCGGTGGCGCGCTAAGCTCAATTCAAAGATTGATGTGACCAAGTGGCGAC 186
Db 44 GlnAlaValAlaAlaSerLysLeuGlyAlaArgValAlaMetIleSerLysLeuGlyAsp 63
QY 187 GATATTTTGGCGCAACACCATTCGTATCTCGAATCCTGGGGATCAATACGATAT 246
Db 64 AspLeuPheAlaGluAsnThrLeuArgAsnPheGluArgPheGlyValAspThrGluHis 83
QY 247 GTAGAAAAGTACCGGTACCAGCGCGGTAGCGCGGATTTTGTCAACCGCACTCC 306
Db 84 ValArgValSerGlyValSerGlyValAlaProIlePheValSerProAspSer 103
QY 307 AGCAACAGATTCGTGATCATCAAGCGCTACAGTCTCTCGCGGGAAGATATCGAT 366
Db 104 ArgAsnArgIleLeuIleValLysGlyAlaAsnArgHisLeuArgProAlaAspIleAsp 123
QY 367 CGCGCGGCGGAAGATTAAATAAATCCAGCTTATTTGTTCTGCAATGGAAGTTCAGCTT 426
Db 124 AlaAlaAlaLysIleGluAlaSerArgLeuValValLeuGlnLeuGluIleAspIle 143
QY 427 GAAACGGTTTATCAGCAATAGATTTGGCAAGAAACACAGGATTAAGTGTATTAAAC 486
Db 144 AspThrValTyThrAlaIleAspPheAlaAlaArgGlyIleProValLeuLeuAsn 163

487 CCTGGCGCAGCATTTACGGGAATTAGATATGCTTATGCTCTAAATGCGATTCTTCTGTA 546
164 ProAlaProGlyValProAspLeuAspPheAlaArgLeuAlaLysLeuGluPheLeuVal 183
547 CCTAATGAACCGAGCTCGGAATATTAAACCGGTATGCCAGTGGATACCTATGACCATATT 606
184 ProAlaGluThrGluLeuAlaLeuValSerGlyMetProThrAspThrProAspAlaVal 203
607 CGCGCAGCGGCAGCTTGGCTGGTGTAGATAAAGGCTGAACAATATTATTGTCCACATGGGC 666
204 GluArgAlaAlaGlySerLeuValGluArgGlyValLysHisValIleValThrLeuGly 223
667 GAGAAAGCGCGCTGTGATCACCGGTGACCGAGGAGTCCATGTTTCCGCGGTTAGAGTG 726
224 GluLysGlySerLeuValSerArgAlaGlyAlaValArgValProProValAlaVal 243
727 AACGCTGTGTATACAGCGCGCGCGCGATGCTTTATCGCTGTTTCCGCGCATTTACTAC 786
244 AspAlaArgAspThrThrGlyAlaGlyAspAlaTyIleGlyCysPheAlaArgHisTy 263
787 GTCCAGACCGGGGATGTGGAAGCGCGCATGAAAGCGGCTCTTTTGGCGCTTTCAGC 846
264 ValAlaThrAlaAspIleProAlaAlaMetArgLeuAlaSerAlaTyAlaAlaHisSer 283
847 GTCACCGGGAAGCGACCCCATCTCTTATCCAGCATTTGAGCAATTAATGAGTATCTT 906
284 ValThrGlyLeuGlyThrGlnLysSerTyAlaAlaAlaAlaThrPheLysLeuVal 303

RESULT 14
Q8U6G0_AGR75 PRELIMINARY; PRT; 312 AA.
ID Q8U6G0_AGR75 PRELIMINARY; PRT; 312 AA.
AC Q8U6G0; Q7CW10;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Ribokinase (AGR_L78p).
GN Name=rbsb; OrderedLocNames=AGR_L78, Atu4847;
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21608550; PubMed=11743193; DOI=10.1126/science.1066804;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Glendinning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A., Gordon D.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Krespan W., Perry M.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens C58.";
RL Science 294:2317-2323 (2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21608551; PubMed=11743194; DOI=10.1126/science.1066803;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M., Mullin L.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmle K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328 (2001).
DR EMBL; AE009413; AAL45641.1; -; Genomic DNA.
DR EMBL; AE008201; AK88602.1; -; Genomic DNA.
DR PIR; AC3153; AC3153.
DR PIR; H98134; H98134.

```
DR HSSP; P05054; IRKD.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0004747; F:ribokinase activity; IEA.
DR GO; GO:0006014; P:D-ribose metabolism; IEA.
DR InterPro; IPR011877; D:ribokin_bact.
DR InterPro; IPR011611; PfkB_region.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; PfkB; 1.
DR PRINTS; PR00990; RIBOKINASE.
DR TIGRFAMs; TIGR02152; D_ribokin_bact; 1.
KW Complete proteome; Kinase.
SQ SEQUENCE 312 AA; 33094 MW; D2A070A3549E35F1 CRC64;

Alignment Scores:
Pred. No.: 312 Length: 312
Score: 878.00 Matches: 168
Percent Similarity: 76.25% Conservative: 60
Best Local Similarity: 56.19% Mismatches: 71
Query Match: 55.22% Indels: 0
DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x Q8U6G0_AGR75 (1-312)
QY 7 ATCCGGGTATTGGCTTAACATGGTGGACCTTATCACCTACACCAACCAGATGCCAAA 66
DB 5 IleGlyValValGlySerAsnMetValAspLeuIleThrTyValAspArgMetProgly 24
QY 67 GAAGGGAACTCGGAAGCCGCGCTTTAAATCGCTCGCGGGAAGAGGGCGAAC 126
DB 25 ProGlyGluThrLeuGluAlaProThrPheGluMetGlyCysGlyGlyLeuAlaAsn 44
QY 127 CAGCGCGGGCGCGCTAACTCAATCAAAAGTATTGATGGTACCACCAAGTGGCGAC 186
DB 45 GlnAlaValAlaAlaAlaArgLeuGlyAlaGluValMetMetValThrArgValGlyAsp 64
QY 187 GATATTTTGGCGCAACACCATTCGTATCTCGAATCTCGGGGATCAATACACGCTAT 246
DB 65 AspValPheAlaAspAsnThrIleArgAsnLeuAlaSerPheGlyValAspThrArgHis 84
QY 247 GTAGAAAAGTACCGTGTACGAGCGCGGTAGCGCGATTTCTGTCGCGGAAGATCCAT 306
DB 85 ValValIysValSerGlyLysSerSerGlyValAlaProIlePheValGluGlnSerGly 104
QY 307 AGCAACACATTCGTATCATCAAGGGCGCTAACAGTTTCTCTGCGCGGAAGATCCAT 366
DB 105 GluAsnSerIleLeuIleValIysGlyAlaAsnAlaAspLeuProValGluValAsp 124
QY 367 CGCGCGCGGAAGATTTAAAAAATGCCAGCTTATTGTTCTGCAACTGGAAAGTTCAGCTT 426
DB 125 LysAlaAlaAlaAspLeuLysGluCysGlyLeuIleLeuMetGlnMetGluValProVal 144
QY 427 GAAACGGTTTATCAGCAATAGAATTTGGCAAGAAACACGGGATTGAAGTGTATTAAAC 486
DB 145 GluThrValTyHisThrIleGluPheAlaAlaGlnAsnGlyIleGluThrIleLeuAsn 164
QY 487 CTGCGCCAGCATTACGGGAATTAGATGCTTATGCTGCTGCTTAATGGATTTCTTTGTA 546
DB 165 ProAlaProAlaAlaAlaAsnLeuAspProGluArgIleArgGlnValThrPheLeuVal 184
QY 547 CTTAATGAAACCGAGCTGGAATATTAAACCGGTATGCCAGTGGATCATCTATGACCATATT 606
DB 185 ProAsnGluSerGluLeuAlaLeuLeuSerGlyLeuProThrAspThrAspGluAspIle 204
QY 607 CGCGCAGCGGCGTCTGCTGCTAGATAAAGGCTGAACAATATTATTGTCCACCATGGGC 666
DB 205 ValArgAlaAlaArgSerLeuIleAlaArgGlyIleArgThrValIleValThrLeuGly 224
QY 667 GAGAAAGCGCGCTGTGATGACCGGTGACAGGAAGTCCATGTTCCGCGCTTTAGAGTG 726
DB 225 AlaArgGlyAlaArgMetIleThrSerAspGluIleValAsnIleGluProValIysVal 244
QY 727 AACGCTGTGTATACAGCGCGCGCGGATGCTTTTATCGGCTGTTTTCGCGCATTTACTAC 786
DB 727 AACGCTGTGTATACAGCGCGCGCGGATGCTTTTATCGGCTGTTTTCGCGCATTTACTAC 786

Db 245 ThrProArgAspThrThrGlyAlaGlyAspAlaPheIleGlySerPheAlaArgPheTyr 264
QY 787 GTCCAGAGCGGGATGTGGAAGCCGCATGAAAAAAGCCGCTCTTTTTCGCGCTTTTCAGC 846
DB 265 AlaGluThrArgGluValValSerSerLeuLysLysAlaSerLeuTyAlaAlaHisSer 284
QY 847 GTCACCGGGAAGACCCCAATCTCTTATCCAGCATTTGAGCAATTTAATGAGTAT 903
DB 285 IleThrArgProGlyThrGlnLysAlaTyAlaSerIleAspGluPheGluAlaPhe 303

RESULT 15
Q6A8F4_PROAC
ID Q6A8F4_PROAC PRELIMINARY; PRT; 304 AA.
AC Q6A8F4;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Sugar kinase, ribokinase family (EC 2.7.1.15).
GN OrderedLocusNames=PPA1211;
OS Propionibacterium acnes.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Propionibacterineae; Propionibacteriaceae; Propionibacterium.
OX NCBI_TaxID=11477;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=KPA171202 / DSM 16379;
RX PubMed=15286373; DOI=10.1126/science.1100330;
RA Brueggemann H., Henne A., Hoster F., Liesegang H., Wiewer A.,
RA Strittmatter A., Hufjer S., Duerre P., Gottschalk G.;
RT "The complete genome sequence of Propionibacterium acnes, a commensal
of human skin.";
RL Science 305:671-673 (2004).
DR EMBL; AE017283; AAT82961.1; -, Genomic DNA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0004747; F:ribokinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006014; P:D-ribose metabolism; IEA.
DR InterPro; IPR011877; D_ribokin_bact.
DR InterPro; IPR011611; PfkB_region.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; PfkB; 1.
DR PRINTS; PR00990; RIBOKINASE.
DR TIGRFAMs; TIGR02152; D_ribokin_bact; 1.
KW Complete proteome; Kinase; Transferase.
SQ SEQUENCE 304 AA; 32566 MW; 3C03C7FA5FAE2080 CRC64;

Alignment Scores:
Pred. No.: 4.19e-64 Length: 304
Score: 831.50 Matches: 168
Percent Similarity: 73.97% Conservative: 48
Best Local Similarity: 57.53% Mismatches: 75
Query Match: 52.30% Indels: 1
DB: 2 Gaps: 1

US-10-049-750-11 (1-921) x Q6A8F4_PROAC (1-304)
QY 1 ATGATATCGCGGTATTGGCTTAACATGGTGGACCTTATCACCTACACCAACCAGATG 60
DB 1 MetAspIleAlaValValGlySerAsnMetValAspLeuIleSerTyIleHisArgMet 20
QY 61 CCCAAAGAAGGGGAACTCTCGAAGCGCGCGCTTTAAATCGGCTCGCGCGGAAAGGG 120
DB 21 ProSerAspGlyGluThrValGluAlaProGluPheArgMetGlyCysGlyLysGly 40
QY 121 GCGAACACAGCGCGTGGCGCGCTAAGCTCAATTCAAAGTATTGATTTGACCAAGTG 180
DB 41 AlaAsnGlnAlaValAlaAlaSerArgLeuGlyAlaGluValMetValThrArgVal 60
QY 181 GCGACGACGATTTTTCGCGCAACACCATTCGTAATCTCGAATCTCGGGGATCAATACG 240
DB 61 GlyAsnAspValPheAlaAspThrThrLeuAspAsnPheArgLysAsnGlyIleAspThr 80
QY 241 ACGTATGTAGAAAAAAGTACCGCTGTACCAAGCGCGCTAGCGCGGATTTTCGTCACGCC 300
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 23, 2005, 22:05:55 ; Search time 90.6485 Seconds
(without alignments)
4682.260 Million cell updates/sec

Title: US-10-049-750-13

Perfect score: 861

Sequence: 1 gtaataaatgcaaaaaa.....acgaagggtgcctatactaa 483

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cpn2.1/USPTO.spool.p/US10049750/runat.23122005.113551.6285/app.query.fasta_1.1742
-DB=A_Geneseq -QFMT=fatan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOCPCL=0 -LOOEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=Blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10049750 @CGN 1 1 476 @runat.23122005.113551.6285 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORE=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A_Geneseq.21.*

1: Genesep1980s.*
2: Genesep1990s.*
3: Genesep2000s.*
4: Genesep2001s.*
5: Genesep2002s.*
6: Genesep2003as.*
7: Genesep2003bs.*
8: Genesep2004s.*
9: Genesep2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	848	98.5	157	4	AAB67589 Amino aci
2	842	97.8	157	8	ADR97103 N-deoxyri
3	728	84.6	158	6	ABP98175 Amino aci
4	719	83.5	159	6	ABP98180 Amino aci
5	637	74.0	158	5	AAG66170 L. helvet
6	629	73.1	133	6	ABP98179 Amino aci
7	405	47.0	84	6	ABP98178 Amino aci
8	209	24.3	168	6	ABP98177 Amino aci
9	166.5	19.3	167	6	ABP98176 Amino aci

10	142	16.5	216	3	AAB11729	Aab11729 Cryptospo
11	142	16.5	216	5	ABJ04047	Abj04047 C parvum
12	142	16.5	1837	3	AAB11726	Aab11726 Cryptospo
13	142	16.5	1837	5	ABJ04044	Abj04044 C parvum
14	138.5	16.1	249	3	AAB11746	Aab11746 C parvum
15	138.5	16.1	249	5	ABJ04059	Abj04059 C parvum
16	138.5	16.1	1721	2	AAB48299	Aa48299 Cryptospo
17	138.5	16.1	1721	2	AAB48299	Aa48299 Cryptospo
18	136.5	16.1	1721	5	ABJ04045	Abj04045 C parvum
19	136.5	15.9	159	5	ABBS3802	Abbs3802 Lactococc
20	134.5	15.6	175	5	ABJ11743	Abj11743 C parvum
21	134.5	15.6	175	5	ABJ04056	Abj04056 C parvum
22	132	15.3	216	3	AAB11734	Aab11734 Cryptospo
23	131	15.2	343	5	ABF69313	Abf69313 Human pol
24	128	14.9	2624	8	ADN97567	Adn97567 Pancreat
25	126.5	14.7	288	2	AAY29082	Aay29082 T. gondii
26	126.5	14.7	288	2	AAY29081	Aay29081 T. gondii
27	126.5	14.7	288	4	AU25553	Aau25553 T. gondii
28	126.5	14.7	288	4	AU25552	Aau25552 T. gondii
29	126.5	14.7	288	7	ADG17391	Adg17391 T. gondii
30	126.5	14.7	288	7	ADG17394	Adg17394 T. gondii
31	126.5	14.7	341	7	ABO73761	Ab073761 Pseudomon
32	126	14.6	1795	4	ABB69806	Abb69806 Drosophil
33	125.5	14.6	605	8	ADS43586	AdS43586 Bacterial
34	124	14.4	292	8	ADX96526	Adx96526 Plant ful
35	124	14.4	580	8	ADN72473	Adn72473 Thale cre
36	120.5	14.0	618	7	ABO79612	Ab079612 Pseudomon
37	119.5	13.9	618	7	ABO74950	Ab074950 Pseudomon
38	119	13.8	281	8	ABO84615	Ab084615 Human can
39	119	13.8	324	8	ADX90939	Adx90939 Plant ful
40	119	13.8	347	8	ADX93820	Adx93820 Plant ful
41	119	13.8	361	8	ADX96938	Adx96938 Plant ful
42	119	13.8	362	8	ADY12362	Ady12362 Plant ful
43	117.5	13.6	333	8	ADY05763	Ady05763 Plant ful
44	117.5	13.6	383	8	ADX90993	Adx90993 Plant ful
45	117.5	13.6	398	8	ADX77255	Adx77255 Plant ful

ALIGNMENTS

RESULT 1

AAB67589

ID AAB67589 standard; protein; 157 AA.

XX

AC AAB67589;

XX

DT 29-MAY-2001 (first entry)

XX

DE Amino acid sequence of a nucleoside 2-deoxyribosyltransferase.

XX

KW Deoxyribonucleoside; deoxyribose 1-phosphate; thymidine phosphorylase;

KW purine nucleoside phosphorylase; phosphopentose mutase;

KW phosphopentose aldolase; fructose 1,6-diphosphate aldolase;

KW deoxyribokinase; nucleoside 2-deoxyribosyltransferase.

XX

OS Lactobacillus leichmannii.

XX

PN WO200114566-A2.

XX

PD 01-MAR-2001.

XX

PF 18-AUG-2000; 2000WO-EP008088.

XX

PR 20-AUG-1999; 99EP-00116425.

XX

PA (HOFF) ROCHE DIAGNOSTICS GMBH.

XX

PA (INSP) INST PASTEUR.

XX

PA (PHAR-) PHARMA-WALDHOFF GMBH & CO KG.

XX

PI Tischer W, Ihlenfeldt H, Barzu O, Sakamoto H, Pistotnik B;

XX

PI Marliere P, Pochet S;

DR WPI; 2001-235026/24.

DR N-PSDB; AAF55445.
 XX In vitro enzymatic synthesis of deoxyribonucleosides comprises reacting
 PT deoxyribose 1-phosphate and a nucleobase to form a deoxyribonucleoside
 PT and an inorganic phosphate.
 XX
 XX Disclosure; Page 62-63; 73pp; English.
 XX
 CC The present sequence represents a nucleoside 2-deoxyribosyltransferase
 CC enzyme. This enzyme is involved in the biosynthesis of
 CC deoxyribonucleosides, and is used in the method of the invention. The
 CC specification describes a method for the in vitro enzymatic synthesis of
 CC deoxyribonucleosides. The method comprises reacting deoxyribose 1-
 CC phosphate and a nucleobase to form a deoxyribonucleoside and an inorganic
 CC phosphate. Enzymes which may be used in the method of the invention
 CC include thymidine phosphorylase, purine nucleoside phosphorylase,
 CC phosphopentose mutase, phosphopentose aldolase, fructose 1,6-diphosphate
 CC aldolase, deoxyribokinase, and nucleoside 2-deoxyribosyltransferase
 XX
 SQ Sequence 157 AA;

Alignment Scores:
 Pred. No.: 3,11e-80 Length: 157
 Score: 848.00 Matches: 157
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 98.49% Indels: 0
 DB: 4 Gaps: 0

US-10-049-750-13 (1-483) x AAB67589 (1-157)

QY 10 ATGCCAAAAAGACGATCTACTTCGGTGGCGGTGGTCTCACTGACCGCCAAACAAAGCC 69
 DB 1 MetProLysLysThrIleTyrPheGlyAlaGlyTrpPheThrAspArgGlnAsnLysAla 20
 QY 70 TACAAGGAGCCATGAAGCCCTCAAGGAAAAACCCAAACGATTGACCTGGAAAAACAGCTAC 129
 DB 21 TyrLysGluAlaMetGluAlaLeuLysGluAsnProThrIleAspLeuGluAsnSerTyr 40
 QY 130 GTTCCCTCGGACCAACAGTACAGGGTATCCGGTTTGATGAACACCCGGAATACCTGCAT 189
 DB 41 ValProLeuAspAsnGlnTyrLysGlyIleArgValAspGluHisProGluTyrLeuHis 60
 QY 190 GACAAGGTTGGCTAGCGCCACCTACAAACAGCTTGAAACGGGATCAAGACCAACGAC 249
 DB 61 AspLysValTrpAlaThrAlaThrTyrAsnAsnAspLeuAsnGlyIleLysThrAsnAsp 80
 QY 250 ATCATGCTGGTGTCTACATCCCTGACGAAGAGCTCGGCTGGGCATGGAACTGGGT 309
 DB 81 IleMetLeuGlyValTyrIleProAspGluGluAspValGlyLeuGlyMetGluLeuGly 100
 QY 310 TAGCCCTTGAGCAAGCAAGTACGTCCTTTTGGTCATCCCGGACGAAGACTACGCAAG 369
 DB 101 TyrAlaLeuSerGlnGlyLysTyrValLeuLeuValIleProAspGluAspTyrGlyLys 120
 QY 370 CCGATCAACCTCATAGCTGGCGGTGAGCGCAACAGCTTGATCAAGATGACCGAGCTGAAG 429
 DB 121 ProIleAsnLeuMetSerTrpGlyValSerAspAsnValIleLysMetSerGlnLeuLys 140
 QY 430 GACTTCAACTTCAACAGCCGCTTCGACTTCTACGAAGTCCCGGTATAC 480
 DB 141 AspPheAsnPheAsnLysProArgPheAspPheTyrGluGlyAlaValTyr 157

RESULT 2
 ADR97103
 ID ADR97103 standard; protein; 157 AA.
 XX
 AC ADR97103;
 XX
 DT 16-DEC-2004 (first entry)
 XX
 DE N-deoxyribosyltransferase (DTP) with G9S mutant, seq id 1.
 XX

KW Cytostatic; antimicrobial; N-dideoxyribosyltransferase; DDTp;
 KW N-deoxyribosyltransferase; DTP; antitumour; infection; ddc; ddi; muten.
 XX
 OS Lactobacillus leichmannii.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 9 /note= "Wild-type Gly replaced with Ser"
 FT
 XX FR2852968-Al.
 XX 01-OCT-2004.
 XX 28-MAR-2003; 2003FR-00003910.
 XX 28-MAR-2003; 2003FR-00003910.
 XX (INSP) INST PASTEUR.
 XX (CNRS) CNRS CENT NAT RECH SCI.
 XX
 XX Kaminski PA;
 XX
 XX WPI; 2004-693069/68.
 DR N-PSDB; ADR97104.
 XX
 XX Method for evolution of proteins to modify characteristics, useful
 PT particularly for evolving an N-deoxyribosyltransferase to produce an N-
 PT dideoxyribosyltransferase for preparation of antitumor nucleosides and
 PT nucleotides.
 XX
 XX Claim 15; SEQ ID NO 2; 51pp; French.
 PS
 XX The invention relates to a method for evolution of a protein (X) so as to
 CC modify its characteristics. Further disclosed are mutated proteins (X*)
 CC produced by the new method and with altered activity, relative to (X).
 CC Also disclosed is a nucleic acid (I) comprising a sequence for N-
 CC dideoxyribosyltransferase (DDTP), obtained from an N-
 CC dideoxyribosyltransferase (DTP) by using the new method. The method is
 CC specifically used to evolve an N-deoxyribosyltransferase (DTP) to an N-
 CC dideoxyribosyltransferase (DDTP), i.e. an enzyme able to transfer
 CC dideoxyribose from one dideoxyribonucleotide to another. Such an enzyme
 CC is used for preparation of nucleotides and nucleosides that have
 CC antitumour activity or can be used to treat infections, particularly ddc
 CC and ddi. The current sequence represents the N-deoxyribosyltransferase
 CC (DTP) amino acid sequence containing a G9S mutation. This protein
 CC displays the activity of N-dideoxyribosyltransferase.
 XX
 XX Sequence 157 AA;

Alignment Scores:
 Pred. No.: 1.33e-79 Length: 157
 Score: 842.00 Matches: 156
 Percent Similarity: 99.36% Conservative: 0
 Best Local Similarity: 99.36% Mismatches: 1
 Query Match: 97.79% Indels: 0
 DB: 8 Gaps: 0

US-10-049-750-13 (1-483) x ADR97103 (1-157)

QY 10 ATGCCAAAAAGACGATCTACTTCGGTGGCGGTGGTCTCACTGACCGCCAAACAAAGCC 69
 DB 1 MetProLysLysThrIleTyrPheSerAlaGlyTrpPheThrAspArgGlnAsnLysAla 20
 QY 70 TACAAGGAGCCATGAAGCCCTCAAGGAAAAACCCAAACGATTGACCTGGAAAAACAGCTAC 129
 DB 21 TyrLysGluAlaMetGluAlaLeuLysGluAsnProThrIleAspLeuGluAsnSerTyr 40
 QY 130 GTTCCCTCGGACCAACAGTACAGGGTATCCGGTTTGATGAACACCCGGAATACCTGCAT 189
 DB 41 ValProLeuAspAsnGlnTyrLysGlyIleArgValAspGluHisProGluTyrLeuHis 60
 QY 190 GACAAGGTTGGGCTACGGCCACCTACAAACAGCTTGAACGGGATCAAGACCAACGAC 249

Db 61 AspLysValTrpAlaThrAlaThrTyrAsnAsnAspLeuAsnGlyIleLysThrAsnAsp 80
 QY 250 ATCATCTGGGTGCTTACATCTCCGACGAAGAAGAGCTCGGCTGGGCATGGAACCTGGGT 309
 Db 81 IleMetLeuGlyValTyrIleProAspGluGluAspValGlyLeuGlyMetGluLeuGly 100
 QY 310 TAGCCCTTACGCAAGGAGTACGCTCTTTGGTTCATCCCGGACGAAGACTACGCGAAG 369
 Db 101 TyrAlaLeuSerGlnGlyLysTyrValLeuLeuValIleProAspGluAspTyrGlyLys 120
 QY 370 CCGATCAACTCATGAGCTGGGCGTCAGCGCAACGTCATCAAGATGAGCCAGCTGAAG 429
 Db 121 ProIleAsnLeuMetSerTrpGlyValSerAspAsnValIleLysMetSerGlnLeuLys 140
 QY 430 GACTTCAACTTCAACAAAGCCGGCTTCGACTTCTACGAAGGTGCCGTATAC 480
 Db 141 AspPheAsnPheAsnLysProArgPheAspPheTyrGluGlyAlaValTyr 157

RESULT 3

ABP98175
 ID ABP98175 standard; protein; 158 AA.

AC ABP98175;

XX 11-AUG-2003 (first entry)

DE Amino acid sequence of N-deoxyribosyltransferase protein.

KW N-deoxyribosyltransferase; ntd gene; enzyme; antitumour chemotherapy;
 KW deoxyribonucleotide synthesis; infection; insecticide; herbicide.

OS Lactobacillus helveticus.

PN WO2003025163-A2.

XX 27-MAR-2003.

XX 12-SEP-2002; 2002WO-FR003120.

XX 14-SEP-2001; 2001FR-00011911.

XX (INSP) INST PASTEUR.

PA (NARE-) INST NAT RECH AGRONOMIQUE.

XX Kaminski P, Tailliez P, Marliere P, Quenee P, Cotaya R;

XX WPI; 2003-313351/30.

DR N-PSDB; ACC43580, ACC43586.

PT New N-deoxyribosyltransferase from Lactobacillus, useful for preparing
 PT deoxyribonucleotides, e.g. for use as antiviral agents, also related
 PT nucleic acid and antibodies.

XX Claim 1; Page 61; 70pp; French.

The present sequence represents a N-deoxyribosyltransferase protein from
 CC Lactobacillus. The enzyme is encoded by the ntd gene. The N-
 CC deoxyribosyltransferase protein is useful to raise specific antibodies,
 CC and for in vitro or in vivo enzymatic synthesis of deoxyribonucleotides,
 CC particularly those containing non-natural bases. These
 CC deoxyribonucleotides are useful for treating bacterial, viral (e.g. HIV),
 CC parasitic or fungal infections, for antitumour chemotherapy, and as
 CC insecticides or herbicides

XX Sequence 158 AA;

Alignment Scores:

Pred. No.:	1,38e-67	Length:	158
Score:	728.00	Matches:	128
Percent Similarity:	95.48%	Conservative:	20
Best Local Similarity:	82.58%	Mismatches:	7
Query Match:	84.55%	Indels:	0

DB: 6 Gaps: 0

US-10-049-750-13 (1-483) x ABP98175 (1-158)

QY 16 AAAAAGACGATCTACTTCGGTCCGCTGGTTCACTGACCGCAAAAACAAAGCTTACAAG 75
 Db 4 LysLysThrLeuTyrPheGlyAlaGlyTrpPheAsnGluLysGlnAsnLysAlaTyrLys 23
 QY 76 GAAGCATGGAAGCCCTCAAGGAAACCAACGATGACCTGGAAACAGTACGTTCCTCC 135
 Db 24 GluAlaMetAlaAlaLeuLysGluAsnProThrValAspLeuGluAsnSerTyrValPro 43
 QY 136 CTGGACACACCTACAAAGGTATCCGGTTGATGACACCCCGGATACCTCGATGACAAG 195
 Db 44 LeuGluAsnGlnTyrLysGlyIleArgIleAspGluHisProGluTyrLeuHisAsnIle 63
 QY 196 GTTTCGGGTACCGCCACCTACAAACGACTTGAACGGGATCAAGCAACGACATCATG 255
 Db 64 GluTrpAlaSerAlaThrTyrHisAsnAspLeuValGlyIleLysThrSerAspValMet 83
 QY 256 CTGGGTGCTACATCCCTGACGAAGACGTCGGCTGGGCATGGAACCTGGGTACGCC 315
 Db 84 LeuGlyValTyrLeuProGluGluAspValGlyLeuGlyMetGluLeuGlyTyrAla 103
 QY 316 TTGAGCCAGGCAAGTACGTCCTTTTGGTCATCCCGGACGAAGACTACGCGAAGCCGATC 375
 Db 104 LeuSerGlnGlyLysTyrIleLeuLeuValIleProAspGluAspTyrGlyLysProIle 123
 QY 376 AACCTCATGAGCTGGGGCTCAGCGACACGTCGATCAAGATGAGCCAGCTGAAGGACTTC 435
 Db 124 AsnLeuMetSerTrpGlyValCysAspAsnAlaIleLysIleSerGluLeuLysAspPhe 143
 QY 436 AACTTCAACAGCCCGCTTCGACTTCTACGAAGGTGCCGTATAC 480
 Db 144 AspPheAsnLysProArgTyrAsnPheTyrAspGlyAlaValTyr 158

RESULT 4

ABP98180
 ID ABP98180 standard; protein; 159 AA.

AC ABP98180;

XX 11-AUG-2003 (first entry)

DE Amino acid sequence of N-deoxyribosyltransferase protein.

KW N-deoxyribosyltransferase; ntd gene; enzyme; antitumour chemotherapy;
 KW deoxyribonucleotide synthesis; infection; insecticide; herbicide.

OS Lactobacillus acidophilus.

PN WO2003025163-A2.

XX 27-MAR-2003.

XX 12-SEP-2002; 2002WO-FR003120.

XX 14-SEP-2001; 2001FR-00011911.

XX (INSP) INST PASTEUR.

PA (NARE-) INST NAT RECH AGRONOMIQUE.

XX Kaminski P, Tailliez P, Marliere P, Quenee P, Cotaya R;

XX WPI; 2003-313351/30.

DR N-PSDB; ACC43585.

PT New N-deoxyribosyltransferase from Lactobacillus, useful for preparing
 PT deoxyribonucleotides, e.g. for use as antiviral agents, also related
 PT nucleic acid and antibodies.

XX Claim 1; Page 68; 70pp; French.

KW N-deoxyribosyltransferase; ntd gene; enzyme; antitumour chemotherapy;
XX deoxyribonucleotide synthesis; infection; insecticide; herbicide.
OS Lactobacillus amylovorus.
XX
XX WO2003025163-A2.
XX
XX PD 27-MAR-2003.
XX
XX PF 12-SEP-2002; 2002WO-FR003120.
XX
XX PR 14-SEP-2001; 2001PR-00011911.
XX
XX (INSP) INST PASTEUR.
PA (NARE-) INST NAT RECH AGRONOMIQUE.
XX
XX Kaminski P, Tailliez P, Marliere P, Quenee P, Cotaya R;
XX WPI; 2003-313351/30.
DR N-PSDB; ACC43584.
XX
XX New N-deoxyribosyltransferase from Lactobacillus, useful for preparing
PT deoxyribonucleotides, e.g. for use as antiviral agents, also related
PT nucleic acid and antibodies.
XX
XX Claim 1; Page 66-67; 70pp; French.
XX
XX The present sequence represents a N-deoxyribosyltransferase protein from
CC Lactobacillus. The enzyme is encoded by the ntd gene. The N-
CC deoxyribosyltransferase protein is useful to raise specific antibodies,
CC and for in vitro or in vivo enzymatic synthesis of deoxyribonucleotides,
CC particularly those containing non-natural bases. These
CC deoxyribonucleotides are useful for treating bacterial, viral (e.g. HIV),
CC parasitic or fungal infections, for antitumour chemotherapy, and as
XX insecticides or herbicides
XX
XX Sequence 133 AA;
Alignment Scores:
Pred. No.: 3,54e-57 Length: 133
Score: 629.00 Matches: 112
Percent Similarity: 96.99% Conservative: 17
Best Local Similarity: 84.21% Mismatches: 4
Query Match: 73.05% Indels: 0
DB: 6 Gaps: 0
US-10-049-750-13 (1-483) x ABP98179 (1-133)
QY 82 ATGGAAGCCCTCAGGAAACCCACGATGACCTGGAAACAGCTACGTTCCCTGGAC 141
Db 1 MetGluAlaLeuLysLysAsnProThrValAspLeuGluAsnSerTyrValProLeuAsp 20
QY 142 ACCAGTACAGGGTATCCGGTTGATGACACCCGGATACCTGCATGACAGGTTGG 201
Db 21 AsnGlnTyrLysGlyIleArgValAspGluHisProGluTyrLeuHisAspIleGluTrp 40
QY 202 GTTACGCCACCTTACCAACACGACTTGAACGGGATCAAGACCAACGACATCATGCTGGGT 261
Db 41 AlaSerSerThrTyrHisAsnAspLeuValGlyIleLysSerSerAspValMetLeuGly 60
QY 262 GTCTACATCCCTGACGAAGAGAGCTCGGCTGGGATGGAACCTGGGTTACGCTTGAC 321
Db 61 ValTyrLeuProGluGluAspValGlyLeuGlyMetGluLeuGlyTyrAlaLeuSer 80
QY 322 CAAAGCAAGTACGCTCTTTTGGTCATCCCGGAGAGACTACGGCAAGCGGATCAACCTC 381
Db 81 GlnGlyLysTyrIleLeuLeuValIleProAspGluAspTyrGlyLysProIleAsnLeu 100
QY 382 ATGAGCTGGGGGTGAGCAGCAACGCTGATCAAGATGAGCAGCTGAAGGACTTCAACTTC 441
Db 101 MetSerTrpGlyValCysAspAsnValIleLysIleSerGluLeuLysAspPhePhe 120
QY 442 AACAGCCCGCTTCGACTTCTACGAAGGTGCGGTATAC 480

Db 121 AsnArgProArgPheAsnPheTyrAspGlyAlaValTyr 133
RESULT 7
ABP98178
ID ABP98178 standard; protein; 84 AA.
XX
XX AC ABP98178;
XX
XX DT 11-AUG-2003 (first entry)
XX
XX Amino acid sequence of N-deoxyribosyltransferase protein.
XX
XX N-deoxyribosyltransferase; ntd gene; enzyme; antitumour chemotherapy;
KW deoxyribonucleotide synthesis; infection; insecticide; herbicide.
XX
XX Lactobacillus crispatus.
XX
XX WO2003025163-A2.
XX
XX PD 27-MAR-2003.
XX
XX PF 12-SEP-2002; 2002WO-FR003120.
XX
XX PR 14-SEP-2001; 2001PR-00011911.
XX
XX (INSP) INST PASTEUR.
PA (NARE-) INST NAT RECH AGRONOMIQUE.
XX
XX Kaminski P, Tailliez P, Marliere P, Quenee P, Cotaya R;
DR WPI; 2003-313351/30.
DR N-PSDB; ACC43583.
XX
XX New N-deoxyribosyltransferase from Lactobacillus, useful for preparing
PT deoxyribonucleotides, e.g. for use as antiviral agents, also related
PT nucleic acid and antibodies.
XX
XX Claim 1; Page 65; 70pp; French.
XX
XX The present sequence represents a N-deoxyribosyltransferase protein from
CC Lactobacillus. The enzyme is encoded by the ntd gene. The N-
CC deoxyribosyltransferase protein is useful to raise specific antibodies,
CC and for in vitro or in vivo enzymatic synthesis of deoxyribonucleotides,
CC particularly those containing non-natural bases. These
CC deoxyribonucleotides are useful for treating bacterial, viral (e.g. HIV),
CC parasitic or fungal infections, for antitumour chemotherapy, and as
XX insecticides or herbicides
XX
XX Sequence 84 AA;
Alignment Scores:
Pred. No.: 1,23e-33 Length: 84
Score: 405.00 Matches: 73
Percent Similarity: 96.43% Conservative: 8
Best Local Similarity: 86.90% Mismatches: 3
Query Match: 47.04% Indels: 0
DB: 6 Gaps: 0
US-10-049-750-13 (1-483) x ABP98178 (1-84)
QY 142 AACCAAGTACAGGGTATCCGGTTGATGAAACCCGGATACCTGCATGACAGGTTGG 201
Db 1 AsnGlnTyrLysGlyIleArgValAspGluHisProGluTyrLeuHisAspIleGluTrp 20
QY 202 GTTACGCCACCTTACCAACACGACTTGAACGGGATCAAGACCAACGACATCATGCTGGGT 261
Db 21 AlaSerAlaThrTyrHisAsnAspLeuValGlyIleLysSerSerAspIleMetLeuGly 40
QY 262 GTCTACATCCCTGACGAAGAGAGCTCGGCTGGGATGGAACCTGGGTTACGCTTGAC 321
Db 41 ValTyrLeuProGluGluAspValGlyLeuGlyMetGluLeuGlyTyrAlaLeuSer 60

```

QY 322 CAAGCCAGTACGCTCTTTGGTCATCCGGAGCAAGACTACGGCAAGCCGATCAACCTC 381
Db ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 LysGlyLysTyrIleLeuLeuValIleProAspGluAspTyrGlyLysProIleAsnLeu 80
QY 382 ATGAGCTGGGGC 393
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81 MetSerTrpGly 84

RESULT 8
ABP98177
ID ABP98177 standard; protein; 168 AA.
XX
AC ABP98177;
XX
DT 11-AUG-2003 (first entry)
XX
DE Amino acid sequence of N-deoxyribosyltransferase protein.
XX
KW N-deoxyribosyltransferase; ntd gene; enzyme; antitumour chemotherapy;
KW deoxyribonucleotide synthesis; infection; insecticide; herbicide.
XX
OS Lactobacillus fermentum.
XX
PN WO2003025163-A2.
XX
PD 27-MAR-2003.
XX
PF 12-SEP-2002; 2002WO-FR003120.
XX
PR 14-SEP-2001; 2001FR-00011911.
XX
PA (INSP ) INST PASTEUR.
XX
PA (NARE-) INST NAT RECH AGRONOMIQUE.
XX
PI Kaminski P, Tailliez P, Marliere P, Quenee P, Cotaya R;
XX
XX WPI; 2003-313351/30.
DR N-PSDB; ACC43582.
XX
PT New N-deoxyribosyltransferase from Lactobacillus, useful for preparing
PT deoxyribonucleotides, e.g. for use as antiviral agents, also related
PT nucleic acid and antibodies.
XX
PS Claim 1; Page 64; 70pp; French.
XX
XX The present sequence represents a N-deoxyribosyltransferase protein from
CC Lactobacillus. The enzyme is encoded by the ntd gene. The N-
CC deoxyribosyltransferase protein is useful to raise specific antibodies,
CC and for in vitro or in vivo enzymatic synthesis of deoxyribonucleotides,
CC particularly those containing non-natural bases. These
CC deoxyribonucleotides are useful for treating bacterial, viral (e.g. HIV),
CC parasitic or fungal infections, for antitumour chemotherapy, and as
CC insecticides or herbicides
XX
SQ Sequence 168 AA;

Alignment Scores:
Pred. No.: 7.03e-13 Length: 168
Score: 209.00 Matches: 55
Percent Similarity: 51.66% Conservative: 23
Best Local Similarity: 36.42% Mismatches: 59
Query Match: 24.27% Indels: 14
DB: Gaps: 7

US-10-049-750-13 (1-483) x ABP98177 (1-168)
QY 25 ATCTACTTCGGTCGGCTGGTTCACTGACCGCCCAAAACAAAGCCATCAAGGAGCCATG 84
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
12 IleTyrLeuAlaThrSerPheAsnGluGlnArgAlaArgIleProGlnAlaLeu 31
QY 85 GAAGCCCTCAAGAAACCAACGATTGACCTGGAAACACAGCTACGTTCCCTCGGACAC 144
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
32 AlaGlnLeuGluAlaAsnProThrValGlyValIleHisGln-----ProPheAspPhe 49

```

```

QY 145 CAGTACAAGGGTATCCGGGTTGTGATGACACACCGGAATACCTGTCATGACAAGGTT---TGG 201
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
50 GlnTyrLysAspAlaArgValAspSerAspProAlaGlyValPheGlySerLeuGluTrp 69
QY 202 GCTACGGCCACCTACACAACAGACTTGAACGGGATCAAGACCAACGACATCATGCTGGGT 261
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
70 GlnIleAlaThrTyrAsnAsnAspLeuAsnAlaValGlyThrSerAspValCysValAla 89
QY 262 GTCTACATCCCTGACGAAGACGCTCGCGCTGGCGATGGAAGTACGCTGGTACGCTTGAGC 321
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90 LeuTyrAspMetAspGlnIleAspGluGlyIleCysMetGluIleGlyMetPheValAla 109
QY 322 CAAGGCAAG---TAGTCTCTTTGGTCATCCGGAGCAAGACTACGGCAAG---CCGATC 375
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110 LeuHisLysProIleValLeuLeuProPheThrLysLysAspLysSerAlaTyrGluAla 129
QY 376 AACCTCATG-----AGCTGGGCGTCAGCGCAACAGCTGATCAAGATG 417
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
130 AsnLeuMetLeuAlaArgGlyValThrThrTrp---LeuGluProAsn-----AspPhe 146
QY 418 ACCCAGCTGAGGACTTCAACTTCAACAGCCG 450
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
147 SerProLeuLysAspPheAsnPheAsnHisPro 157

RESULT 9
ABP98176
ID ABP98176 standard; protein; 167 AA.
XX
AC ABP98176;
XX
DT 11-AUG-2003 (first entry)
XX
DE Amino acid sequence of N-deoxyribosyltransferase protein.
XX
KW N-deoxyribosyltransferase; ptd gene; enzyme; antitumour chemotherapy;
KW deoxyribonucleotide synthesis; infection; insecticide; herbicide.
XX
OS Lactobacillus helveticus.
XX
PN WO2003025163-A2.
XX
PD 27-MAR-2003.
XX
PF 12-SEP-2002; 2002WO-FR003120.
XX
PR 14-SEP-2001; 2001FR-00011911.
XX
PA (INSP ) INST PASTEUR.
XX
PA (NARE-) INST NAT RECH AGRONOMIQUE.
XX
PI Kaminski P, Tailliez P, Marliere P, Quenee P, Cotaya R;
XX
XX WPI; 2003-313351/30.
DR N-PSDB; ACC43581.
XX
XX New N-deoxyribosyltransferase from Lactobacillus, useful for preparing
PT deoxyribonucleotides, e.g. for use as antiviral agents, also related
PT nucleic acid and antibodies.
XX
PS Claim 1; Page 62-63; 70pp; French.
XX
XX The present sequence represents a N-deoxyribosyltransferase protein from
CC Lactobacillus. The enzyme is encoded by the ptd gene. The N-
CC deoxyribosyltransferase protein is useful to raise specific antibodies,
CC and for in vitro or in vivo enzymatic synthesis of deoxyribonucleotides,
CC particularly those containing non-natural bases. These
CC deoxyribonucleotides are useful for treating bacterial, viral (e.g. HIV),
CC parasitic or fungal infections, for antitumour chemotherapy, and as
CC insecticides or herbicides
XX
SQ Sequence 167 AA;

```

Alignment Scores:
Pred. No.: 2,11e-08 Length: 167
Score: 166.50 Matches: 49
Percent Similarity: 51.83% Conservative: 36
Best Local Similarity: 29.88% Mismatches: 64
Query Match: 19.34% Indels: 15
DB: 7 Gaps: 7

US-10-049-750-13 (1-483) x ABP98176 (1-167)

QY 10 ATGCAAAAAGACGATCTACTTCGTGGCGCTGGTTCACCTGACCGCCCAAAACAAAGCC 69
DB 5 ValProThrGlyLysIleTyrLeuGlySerProPheTyrSerAspAlaGlnArgGluArg 24
QY 70 TACAAGGAAGCCATGGAAGCCCTCAAGGAAGAAACCCCAACGATTCACCTGGAAAACAGCTAC 129
DB 25 AlaAlaLysAlaLysGluLeuLeuAlaLysAsn-----LeuSerIleAlaHisValPhe 42
QY 130 GTTCCCTGACAAACAGATACAGAGGTATCCGGTGTGATGAACACCCCGGAATAC----- 183
DB 43 PheProPheAspAspGlyPheThrAsp---ProAspGluLysAsnProGluIleGlyGly 61
QY 184 CTGCATGACAAAGCTTGGGCTACGGCCACCTCAACAACAGACTTGAACGGGATCAAGACC 243
DB 62 IleArgSerMetValIrpArgAspAlaThrTyrGlnAsnAspLeuThrGlyIleSerAsn 81
QY 244 AAGCAGCATCATGCTGGGTGTCTACATCCCTGACGAAGAAGACGCTGGCGCTGGGCATGAA 303
DB 82 AlaThrCysGlyValPheLeuTyrAspMetAspGlnLeuAspAspGlySerAlaPheGlu 101
QY 304 CTGGTTACGCTTGGACCAAGCAAGTACGCTCTTTTGGTC-----ATCCCG 351
DB 102 IleGlyPheMetArgAlaMetHisLysProValIleLeuValProPheThrGluHisPro 121
QY 352 GACGAAGACTACGGCAAGCGCATCAACCTCATG---AGCTGGGGCTCAGC----- 399
DB 122 GluLysGlu-----LysLysMetAsnLeuMetIleAlaGlnGlyValThrThrIleIle 139
QY 400 GACAAGCTGATCAAGATGACGACGAGCTGAAGAGCTTCAACTTCAACAAGCGCGCTTCGAC 459
DB 140 AspGlyAsnThrGluPheGluLysLeuAlaAspTyrAsnPheAsnGluCysProPheAsn 159
QY 460 TTCTACGAAGGT 471
DB 160 ProValArgGly 163

RESULT 10
AAB11729
ID AAB11729 standard; protein; 216 AA.
XX
AC AAB11729;
XX
DT 28-OCT-2000 (first entry)
XX
DE Cryptosporidium parvum Iowa isolate GP900, domain 2.
XX
DE GP900; Iowa isolate; glycoprotein; antibody; cryptosporidiosis;
KW competitive inhibition; attachment; invasion; ligand binding; sporozoite;
KW merozoite; diarrhoea; protozoacide; domain 2; mucin-like.
XX
OS Cryptosporidium parvum.
XX
PN US6071518-A.
XX
PN 06-JUN-2000.
XX
PD 12-SBP-1997; 97US-00928361.
XX
PF 29-MAY-1992; 92US-00891301.
XX
PR 01-JUN-1993; 93US-00071880.
XX
PR 03-APR-1995; 95US-00415751.
XX
PR 14-AUG-1996; 96US-00700651.
XX
PR 13-SBP-1996; 96US-0026062P.

XX PA (REGC) UNIV CALIFORNIA.
XX Petersen C;
XX WPI; 2000-422065/36.
XX New GP900 protein fragments and fusion proteins of Cryptosporidium
PT parvum, useful for detecting the presence of the parasite, and diagnosing
PT or treating Cryptosporidium infections by competitive inhibition of the
PT function of GP900.
XX Claim 3; Col 71-72; 59pp; English.
XX The invention relates to the GP900 glycoprotein of the protozoan
CC Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion
CC proteins comprising GP900 fragments. The invention also relates to the
CC administration of GP900 or fragments thereof to a host to elicit anti-
CC GP900 antibody production, and to a method of cryptosporidiosis treatment
CC or prophylaxis comprising administration of anti-GP900 antibodies to an
CC individual. Cryptosporidium parvum GP900 and GP900 fragments are able to
CC competitively inhibit sporozoite or merozoite attachment or invasion, and
CC are also useful for the generation of anti-GP900 antibodies. The
CC antibodies also inhibit sporozoite or merozoite attachment/invasion and
CC additionally inhibit the binding of GP900 ligands to GP900. GP900
CC proteins, fragments and antibodies may therefore be used to treat or
CC prevent cryptosporidiosis. Infection with Cryptosporidium is a common
CC cause of diarrhoea in humans and causes life-threatening diarrhoea in
CC immunocompromised persons. Cryptosporidiosis can be contracted from
CC contaminated municipal water supplies (e.g., public swimming pools). It
CC is also a cause of disease in animals, resulting in financial losses in
CC agriculture. GP900 fragments, fusion proteins and antibodies may also be
CC used for the diagnosis of Cryptosporidium parvum infections, and for the
CC detection of the parasite in the environment. Sequences AAB11728-B11732
CC represent the different domains of the Cryptosporidium parvum Iowa
CC isolate GP900 protein (AAB11726). The present sequence is that of the
CC mucin-like domain 2 of GP900
XX SQ Sequence 216 AA;
Alignment Scores:
Pred. No.: 8.82e-06 Length: 216
Score: 142.00 Matches: 45
Percent Similarity: 46.10% Conservative: 26
Best Local Similarity: 29.22% Mismatches: 67
Query Match: 16.49% Indels: 16
DB: 3 Gaps: 3
US-10-049-750-13 (1-483) x AAB11729 (1-216)
QY 26 TCTACTTCGGTGGCGCTGGTTCACCTGACCGCCCAAAACAAAGCCTCAAGGAAGCCA--- 82
DB 52 Thr 71
QY 83 ---TGAAGCCCTCAAGGAAAAACCAACGATTCACCTGGAAAAACAGCTACGTTCCCTGG 139
DB 72 Thr 91
QY 140 ACAACAGTAGTACAGGTATCCGGTGTGATGAACACCCGGATACCTGCGATGACAAAGTTT 199
DB 92 Thr 111
QY 200 GGGCTACGGCCACTCAACAACAGCTTGAACCGGATCAAGCAACACGATCATGCTGCG 259
DB 112 ThrThrLysLysProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 131
QY 260 GTGTCTACATCCCTGACGAAGAAGACGTGGCGCTGGGCATGGAACCTGGGTACGCTTGA 319
DB 132 Thr 145
QY 320 GCCAAGCAAGTACGTCCTTTTGGTCATCCCGGACGACGACTACGCGAAGCCATCAACC 379
DB 146 LysLysProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 165

Alignment Scores:
Pred. No.: 2,11e-08 Length: 167
Score: 166.50 Matches: 49
Percent Similarity: 51.83% Conservative: 36
Best Local Similarity: 29.88% Mismatches: 64
Query Match: 19.34% Indels: 15
DB: 7

US-10-049-750-13 (1-483) x ABP98176 (1-167)

QY 10 ATGCAAAAAGACGATCTACTTCGCTGCGGCTGCTCACTGACCGCCAAACAAAGCC 69
DB 5 ValProThrGlyLysIleTyrLeuGlySerProPheTyrSerAspAlaGlnArgGluArg 24
QY 70 TACAAGGAAGCCATGGAAGCCCTCAAGGAAGAAACCCCAACGATTGACCTGGAAAACAGCTAC 129
DB 25 AlaAlaLysAlaLysGluLeuLeuAlaLysAsn-----LeuSerIleAlaHisValPhe 42
QY 130 GTTCCCTGACAACACAGTACAGAGGATCCGGGTTGATGCAACACCCGGAATAC----- 183
DB 43 PheProPheAspAspGlyPheThrAsp---ProAspGluLysAsnProGluIleGlyGly 61
QY 184 CTGCGATGACAAGCTTTGGGCTACGGCCACCTCAACAACAGACTTGAAACGGGATCAAGACC 243
DB 62 IleArgSerMetValIrpArgAspAlaThrTyrGlnAsnAspLeuThrGlyIleSerAsn 81
QY 244 AAGCAGCATCATGCTGGGTGTCTACATCCCTGACGAGAAGAGAGCTGGCTGGGCATGGAA 303
DB 82 AlaThrCysGlyValPheLeuTyrAspMetAspGlnLeuAspAspGlySerAlaPheGlu 101
QY 304 CTGGTTTACGCTTGGAGCCAGCAAGTACGCTCTTTTGGTC-----ATCCCG 351
DB 102 IleGlyPheMetArgAlaMetHisLysProValIleLeuValProPheThrGluHisPro 121
QY 352 GACGAAGACTACGGCAGCGCATCAACCTCATG---AGCTGGGGCTCAGC----- 399
DB 122 GluLysGlu-----LysLysMetAsnLeuMetIleAlaGlnGlyValThrThrIleIle 139
QY 400 GACAAGCTGATCAAGATGAGCAGCTGGAAGGACTTCAACTTCAACAAGCGCGCTTCGAC 459
DB 140 AspGlyAsnThrGluPheGluLysLeuAlaAspTyrAsnPheAsnGluCysProPheAsn 159
QY 460 TTCTACGAAGGT 471
DB 160 ProValArgGly 163

RESULT 10
AAB11729
ID AAB11729 standard; protein; 216 AA.
XX
AC AAB11729;
XX
DT 28-OCT-2000 (first entry)
XX
DE Cryptosporidium parvum Iowa isolate GP900, domain 2.
XX
DE GP900; Iowa isolate; glycoprotein; antibody; cryptosporidiosis;
KW competitive inhibition; attachment; invasion; ligand binding; sporozoite;
KW merozoite; diarrhoea; protozoacide; domain 2; mucin-like.
XX
OS Cryptosporidium parvum.
XX
PN US6071518-A.
XX
PN 06-JUN-2000.
XX
PD 12-SBP-1997; 97US-00928361.
XX
PF 29-MAY-1992; 92US-00891301.
XX
PR 01-JUN-1993; 93US-00071880.
XX
PR 03-APR-1995; 95US-00415751.
XX
PR 14-AUG-1996; 96US-00700651.
XX
PR 13-SBP-1996; 96US-0026062P.

XX PA (REGC) UNIV CALIFORNIA.
XX Petersen C;
XX WPI; 2000-422065/36.
XX New GP900 protein fragments and fusion proteins of Cryptosporidium
PT parvum, useful for detecting the presence of the parasite, and diagnosing
PT or treating Cryptosporidium infections by competitive inhibition of the
PT function of GP900.
XX Claim 3; Col 71-72; 59pp; English.
XX
XX The invention relates to the GP900 glycoprotein of the protozoan
CC Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion
CC proteins comprising GP900 fragments. The invention also relates to the
CC administration of GP900 or fragments thereof to a host to elicit anti-
CC GP900 antibody production, and to a method of cryptosporidiosis treatment
CC or prophylaxis comprising administration of anti-GP900 antibodies to an
CC individual. Cryptosporidium parvum GP900 and GP900 fragments are able to
CC competitively inhibit sporozoite or merozoite attachment or invasion, and
CC are also useful for the generation of anti-GP900 antibodies. The
CC antibodies also inhibit sporozoite or merozoite attachment/invasion and
CC additionally inhibit the binding of GP900 ligands to GP900. GP900
CC proteins, fragments and antibodies may therefore be used to treat or
CC prevent cryptosporidiosis. Infection with Cryptosporidium is a common
CC cause of diarrhoea in humans and causes life-threatening diarrhoea in
CC immunocompromised persons. Cryptosporidiosis can be contracted from
CC contaminated municipal water supplies (e.g., public swimming pools). It
CC is also a cause of disease in animals, resulting in financial losses in
CC agriculture. GP900 fragments, fusion proteins and antibodies may also be
CC used for the diagnosis of Cryptosporidium parvum infections, and for the
CC detection of the parasite in the environment. Sequences AAB11728-B11732
CC represent the different domains of the Cryptosporidium parvum Iowa
CC isolate GP900 protein (AAB11726). The present sequence is that of the
CC mucin-like domain 2 of GP900
XX SQ Sequence 216 AA;
Alignment Scores:
Pred. No.: 8.82e-06 Length: 216
Score: 142.00 Matches: 45
Percent Similarity: 46.10% Conservative: 26
Best Local Similarity: 29.22% Mismatches: 67
Query Match: 16.49% Indels: 16
DB: 3 Gaps: 3
US-10-049-750-13 (1-483) x AAB11729 (1-216)
QY 26 TCTACTTCGGTGGCGGCTGGTTCACCTGACCGCCAAACAAAGCCTCAAGGAAGCCA--- 82
DB 52 Thr 71
QY 83 ---TGAAGCCCTCAAGGAAAAACCAACGATTCAGCTGGAAAAACAGCTACGTTCCCTGG 139
DB 72 Thr 91
QY 140 ACAACAGTAGTACAGGATCCGGGTGTGATGAACACCCGGATACCTGCGATGACGAGTTT 199
DB 92 Thr 111
QY 200 GGGCTAGCGGCACCTACAACAACGACTTGAAACGGGATCAAGCAACACGATCATGCTGG 259
DB 112 ThrThrLysLysProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 131
QY 260 GTGTCTACATCCCTGACGAAGAAGACGCTCGGCTGGGCATGGAACTGGGTTACGCTTGA 319
DB 132 Thr 145
QY 320 GCCAAGCAAGTACGCTCTTTTGGTCATCCCGGACGACGACTACGCAAGCCCATCAACC 379
DB 146 LysLysProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 165


```

Db 474 ThrThrThrThrThrThrThrThrThr-----ThrThrThr 485
Qy 440 TCAACAGCCGGCTTCGACTCTACGAGGTGCGTATACT 481
Db 486 ThrThrThrThrAlaThrThrThrThrThrLysLysProThrThr 499

RESULT 14
AAB11746
ID AAB11746 standard; protein; 249 AA.
XX
AC AAB11746;
XX
DT 28-OCT-2000 (first entry)
XX
DE C. parvum NINC isolate GP900 variant domain 2, SEQ ID NO:20.
XX
KW GP900; NINC isolate; glycoprotein; antibody; cryptosporidiosis;
KW competitive inhibition; attachment; invasion; ligand binding; sporozoite;
KW merozoite; diarrhoea; protozoacide; domain 2; mucin-like; variant.
XX
OS Cryptosporidium parvum.
XX
PN US6071518-A.
XX
PD 06-JUN-2000.
XX
PF 12-SEP-1997; 97US-00928361.
XX
PR 29-MAY-1992; 92US-00891301.
PR 01-JUN-1993; 93US-00071880.
PR 03-APR-1995; 95US-00415751.
PR 14-AUG-1996; 96US-00700651.
PR 13-SEP-1996; 96US-0026062P.
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Petersen C;
XX
DR WPI; 2000-422065/36.
XX
PT New GP900 protein fragments and fusion proteins of Cryptosporidium
PT parvum, useful for detecting the presence of the parasite, and diagnosing
PT or treating Cryptosporidium infections by competitive inhibition of the
PT function of GP900.
XX
PS Claim 4; Col 17-18; 59pp; English.
XX
CC The invention relates to the GP900 glycoprotein of the protozoan
CC Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion
CC proteins comprising GP900 fragments. The invention also relates to the
CC administration of GP900 or fragments thereof to a host to elicit anti-
CC GP900 antibody production, and to a method of cryptosporidiosis treatment
CC or prophylaxis comprising administration of anti-GP900 antibodies to an
CC individual. Cryptosporidium parvum GP900 and GP900 fragments are able to
CC competitively inhibit sporozoite or merozoite attachment or invasion, and
CC are also useful for the generation of anti-GP900 antibodies. The
CC antibodies also inhibit sporozoite or merozoite attachment/invasion and
CC additionally inhibit the binding of GP900 ligands to GP900. GP900
CC proteins, fragments and antibodies may therefore be used to treat or
CC prevent cryptosporidiosis. Infection with Cryptosporidium is a common
CC cause of diarrhoea in humans and causes life-threatening diarrhoea in
CC immunocompromised persons. Cryptosporidiosis can be contracted from
CC contaminated municipal water supplies (e.g., public swimming pools). It
CC is also a cause of disease in animals, resulting in financial losses in
CC agriculture. GP900 fragments, fusion proteins and antibodies may also be
CC used for the diagnosis of Cryptosporidium parvum infections, and for the
CC detection of the parasite in the environment. The present sequence
CC represents a variant of the mucin-like domain 2 (AAB11734) of the
CC Cryptosporidium parvum NINC isolate GP900 protein (AAB11727)
XX
SQ Sequence 249 AA;
XX
Alignment Scores:

```

```

Pred. No.: 2.16e-05 Length: 249
Score: 138.50 Matches: 44
Percent Similarity: 43.75% Conservative: 26
Best Local Similarity: 27.50% Mismatches: 89
Query Match: 16.09% Indels: 1
DB: 3 Gaps: 1

US-10-049-750-13 (1-483) x AAB11746 (1-249)
Qy 2 TATACTAAATGCCAAAGAGACGATCTACTTCGGTGGCGGTTCACGACCGCCAAA 61
Db 9 TyrThrLysCysValGlyValLysHisThrThrThrThrThrThrThrThrThr 28
Qy 62 ACAAGAGCCCTACAAGAGAGCCATGGAAGCCCTCAAGGAAACCAACGATGACCTGGAAA 121
Db 29 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 48
Qy 122 ACAGCTAGTTCCCTCGGACAAACAGTACAAGGGTATCCGGGTTCATGAACACCGCGAAT 181
Db 49 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 68
Qy 182 ACCTGCATGACAAAGGTTTGGGCTAGGGCCACTACAACAACGACTTGAAACGGGATCAAAGA 241
Db 69 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 87
Qy 242 CCAACGACATCATGCTGGGTGTCTACATCCCTGACGAGAAGACGTCGCGCTGGGCATGG 301
Db 88 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 107
Qy 302 AACTGGGTTCAGCCCTTGAGCCCAAGCAAGTACGTCCTTTGGTCATCCGGAGCAAGACT 361
Db 108 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 127
Qy 362 ACGGCAAGCCGATCAACCTCATGAGCTGGGCGCTCAGCAACACGTCGATCAAGATGAGCC 421
Db 128 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 147
Qy 422 AGCTGAAGGACTTCAACTTCAACAGCCGCGCTTCGACTTCTACGAAGTGCCGTATACT 481
Db 148 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 167

RESULT 15
ABJ04059
ID ABJ04059 standard; protein; 249 AA.
XX
AC ABJ04059;
XX
DT 27-SEP-2002 (first entry)
XX
DE C parvum GP900 protein fragment SEQ ID NO: 20.
XX
KW Cryptosporidium detection; GP900; P68; cryptopain; cryptosporidiosis.
XX
OS Cryptosporidium parvum.
XX
PN WO200194631-A1.
XX
PD 13-DEC-2001.
XX
PF 14-MAY-2001; 2001WO-US015624.
XX
PR 06-JUN-2000; 2000US-00588995.
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Petersen C, Barnes DA, Nelson RG, Gut J;
XX
DR WPI; 2002-566447/60.
XX
PT Detecting Cryptosporidium in biological and environmental samples and
PT diagnosis of cryptosporidiosis involves, contacting the sample with
PT Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or RNA.
XX
XX

```

PS Disclosure; Page 128-129; 157pp; English.

XX The present invention relates to a method of detecting Cryptosporidium in
CC biological and environmental samples, and of diagnosing
CC cryptosporidiosis. This involves obtaining a sample and contacting it
CC with Cryptosporidium GP900, P68 or cryptosporin antigen, antibody, DNA or
CC RNA, or its variant, mutant or fragment. The method is also useful for
CC detecting and identifying individual Cryptosporidium isolates based on
CC the genetic characteristics, and for diagnosis of prior or concurrent
CC Cryptosporidium infection. The present sequence is a C. parvum protein
CC sequence used in the exemplification of the invention

XX Sequence 249 AA;

Alignment Scores:

Pred. No.:	2.16e-05	Length:	249
Score:	138.50	Matches:	44
Percent Similarity:	43.75%	Conservative:	26
Best Local Similarity:	27.50%	Mismatches:	89
Query Match:	16.09%	Indels:	1
DB:	5	Gaps:	1

US-10-049-750-13 (1-483) x ABJ04059 (1-249)

QY	2	TATACTAAATGCAAAAAGACGATCTACTCGTGCCGGCTGGTTCACTGACCGCAAA	61
DB	9	TyrThrLysCysValGlyVallyHisThrThrThrThrThrThrThrThrThrThr	28
QY	62	ACAAGCCTACAGGAGCCATGGAGCCCTCAAGGAAACCAACGATTGACCTGGAA	121
DB	29	ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr	48
QY	122	ACAGCTACCTGCCCTGGACACACAGTACAGGGTATCCGGGTGTGATGAACACCGGAAT	181
DB	49	ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr	68
QY	182	ACCTGCATGACAAAGTTTGGGCTACGCCACCTACAAACGACTTGAACGGGATCAAGA	241
DB	69	ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr	87
QY	242	CCAAAGCATCATGCTGGGTGTACATCCCTGACGAGAGACGTCGGCCTGGGCATGG	301
DB	88	ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr	107
QY	302	AACTGGGTTACGCCTTGAGCCCAAGCAAGTACGTCTTTTGGTCTCCCGACGAAGACT	361
DB	108	ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr	127
QY	362	ACGGCAAGCCGATCAACCTCATGAGCTGGGGGTACGCGACAAACGTGATCAAGATGAGCC	421
DB	128	ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr	147
QY	422	ACGTGAGGACTTCACTTCAACAGCCGGCTTCGACTTCTACGAGGTGCGGTACT	481
DB	148	ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr	167

Search completed: December 23, 2005, 22:29:56
Job time : 96.6485 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 23, 2005, 22:17:41 ; Search time 19.609 Seconds
(without alignments)
4072.865 Million cell updates/sec

Title: US-10-049-750-13

Perfect score: 861

Sequence: 1 gatactaataatgccaaaaa.....acgaaggcgctatactaa 483

Scoring table:

BLASUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10049750/runat_23122005_113553_6317/app_query_fasta_1.1742
-DB=Issued_Patents_AA -OFMT=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10049750@cgn1.1.101@runat_23122005_113553_6317 -NCPU=6 -ICPU=3
-NM MAP -LARGEQUERY -NEG SCORES=0 -WAIT DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5 COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/6 COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/H COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/PTCUS COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/RE COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	142	16.5	216	2	US-08-928-361B-8
2	142	16.5	216	2	US-08-928-361B-8
3	142	16.5	1837	2	US-08-928-361B-5
4	142	16.5	1837	2	US-08-928-361B-5
5	138.5	16.1	249	2	US-08-700-651-15
6	138.5	16.1	249	2	US-08-928-361B-20
7	138.5	16.1	249	2	US-08-928-361B-20
8	138.5	16.1	1721	2	US-08-700-651-5
9	138.5	16.1	1721	2	US-08-928-361B-6
10	138.5	16.1	1721	2	US-08-928-361B-6
11	134.5	15.6	175	2	US-08-700-651-12
12	134.5	15.6	175	2	US-08-928-361B-17

13	134.5	15.6	175	2	US-09-588-995A-17	Sequence 17, Appl
14	132	15.3	216	2	US-08-928-361B-27	Sequence 27, Appl
15	126.5	14.7	288	2	US-09-216-393B-341	Sequence 341, App
16	126.5	14.7	288	2	US-09-216-393B-344	Sequence 344, App
17	126.5	14.7	341	2	US-09-252-991A-22507	Sequence 22507, A
18	125.5	14.6	605	2	US-09-487-558B-428	Sequence 428, App
19	123	14.3	162	2	US-08-700-651-13	Sequence 13, Appl
20	120.5	14.0	618	2	US-09-252-991A-28358	Sequence 28358, A
21	119.5	13.9	618	2	US-09-252-991A-23696	Sequence 23696, A
22	117	13.6	1113	2	US-09-252-991A-29215	Sequence 29215, A
23	113	13.1	138	2	US-08-700-651-10	Sequence 10, Appl
24	113	13.1	138	2	US-08-928-361B-15	Sequence 15, Appl
25	113	13.1	138	2	US-09-588-995A-15	Sequence 15, Appl
26	113	13.1	181	2	US-09-252-991A-21966	Sequence 21966, A
27	112.5	13.1	150	2	US-08-928-361B-18	Sequence 18, Appl
28	112.5	13.1	150	2	US-09-588-995A-18	Sequence 18, Appl
29	112.5	13.1	658	2	US-09-252-991A-25861	Sequence 25861, A
30	111.5	13.0	611	2	US-09-252-991A-17597	Sequence 17597, A
31	110.5	12.8	406	2	US-09-252-991A-19857	Sequence 19857, A
32	109.5	12.7	508	2	US-09-252-991A-25549	Sequence 25549, A
33	109	12.7	130	2	US-08-700-651-8	Sequence 8, Appl
34	109	12.7	130	2	US-08-928-361B-13	Sequence 13, Appl
35	109	12.7	130	2	US-09-588-995A-13	Sequence 13, Appl
36	108.5	12.6	154	2	US-09-252-991A-32553	Sequence 32553, A
37	108.5	12.6	174	2	US-09-252-991A-27275	Sequence 27275, A
38	108.5	12.6	2870	2	US-09-479-467A-15	Sequence 15, Appl
39	108.5	12.6	2870	2	US-09-655-160-15	Sequence 15, Appl
40	108.5	12.6	3178	2	US-09-479-467A-4	Sequence 4, Appl
41	108.5	12.6	3178	2	US-09-655-160-4	Sequence 4, Appl
42	108	12.5	130	2	US-08-700-651-9	Sequence 9, Appl
43	108	12.5	130	2	US-08-928-361B-14	Sequence 14, Appl
44	108	12.5	130	2	US-09-588-995A-14	Sequence 14, Appl
45	108	12.5	453	2	US-09-686-583B-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1

US-08-928-361B-8

; Sequence 8, Application US/08928361B

; Patent No. 6071518

; GENERAL INFORMATION:

; APPLICANT: Petersen, Carolyn

; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,

; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS

; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM

; TITLE OF INVENTION: SPECIES INFECTIONS

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PETERS, VERNY, JONES & BIKSA

; STREET: 385 Sherman Avenue, Suite 6

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94306-1840

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatenIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/928,361B

; FILING DATE: 12-SEP-1997

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/026,062

; FILING DATE: 13-SEP-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: VERNY, Hana

; REGISTRATION NUMBER: 30,518

; REFERENCE/DOCKET NUMBER: 480.76-1 (HV)

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-324-1677

APPLICANT: LEECH, JAMES
APPLICANT: NELSON, RICHARD, C.
APPLICANT: GUT, JIRI
TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
TITLE OF INVENTION: INFECTIONS
FILE REFERENCE: 480.19-4(HV)
CURRENT APPLICATION NUMBER: US/08/700,651B
CURRENT FILING DATE: 1997-08-14
EARLIER APPLICATION NUMBER: 08/415,751
EARLIER FILING DATE: 1995-04-03
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 249
TYPE: PRT
ORGANISM: Cryptosporidium parvum
FEATURE:
OTHER INFORMATION: mutant/variant of SEQ ID NO:5
US-08-700-651-15

Alignment Scores:
Pred. No.: 5,22e-07 Length: 249
Score: 138.50 Matches: 44
Percent Similarity: 43.75% Conservative: 26
Best Local Similarity: 27.50% Mismatches: 89
Query Match: 16.09% Indels: 1
DB: 2 Gaps: 1

US-10-049-750-13 (1-483) x US-08-700-651-15 (1-249)

QY 2 TATACATAAATGCAAAAAGACGATCTACTTCGGTCCGGCTGGTTCACGTACGCCCAAA 61
Db 9 TyrThrLysCysValGlyVallYshsThrThrThrThrThrThrThrThrThrThrThr 28
QY 62 ACAAGAGCTTACAGGAGCCATGGAAGCCCTCAAGGAAACCCACGATTCACCTGGAAA 121
Db 29 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 48
QY 122 ACAGCTAGCTTCCCTGGACACCAAGGATACAGGGTATCCGGTTCGATGAACACCGGAAT 181
Db 49 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 68
QY 182 ACCTGCATGACAAAGTTGGGCTACGCCACCTACACAGCACTTGACGGGATCAAGA 241
Db 69 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 87
QY 242 CCAAGCATCATGCTGGGTGCTACATCCCTGACGAGAACGCTCGGCTGGGCTGG 301
Db 88 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 107
QY 302 AACTGGGTACGCTTCGACCAAGGACGATGATCGGTTCATCCGGACGAGACT 361
Db 108 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 127
QY 362 ACGCAAGCCGATCAACTCATGAGCTGGGGCTGACGACAAACGATGATCAAGATGACC 421
Db 128 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 147
QY 422 ACCTGAGGACTTCACCTCAACAGCCGGCTTCGACTTCACGAGGTGCGGTATAC 481
Db 148 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 167

RESULT 6

US-08-928-361B-20
Sequence 20, Application US/08928361B
Patent No. 6071518
GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
TITLE OF INVENTION: SPECIES INFECTIONS

NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verny, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 249 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-928-361B-20

Alignment Scores:
Pred. No.: 5,22e-07 Length: 249
Score: 138.50 Matches: 44
Percent Similarity: 43.75% Conservative: 26
Best Local Similarity: 27.50% Mismatches: 89
Query Match: 16.09% Indels: 1
DB: 2 Gaps: 1

US-10-049-750-13 (1-483) x US-08-928-361B-20 (1-249)

QY 2 TATACATAAATGCAAAAAGACGATCTACTTCGGTCCGGCTGGTTCACGTACGCCCAAA 61
Db 9 TyrThrLysCysValGlyVallYshsThrThrThrThrThrThrThrThrThrThrThr 28
QY 62 ACAAGAGCTTACAGGAGCCATGGAAGCCCTCAAGGAAACCCACGATTCACCTGGAAA 121
Db 29 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 48
QY 122 ACAGCTAGCTTCCCTGGACACCAAGGATACAGGGTATCCGGTTCGATGAACACCGGAAT 181
Db 49 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 68
QY 182 ACCTGCATGACAAAGTTGGGCTACGCCACCTACAAACGACCTTGACGGGATCAAGA 241
Db 69 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 87
QY 242 CCAAGCATCATGCTGGGTGCTACATCCCTGACGAGAACGCTCGGCTGGGCTGG 301
Db 88 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 107
QY 302 AACTGGGTACGCTTCGACCAAGGACGATGATCGGTTCATCCGGACGAGACT 361
Db 108 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 127
QY 362 ACGCAAGCCGATCAACTCATGAGCTGGGGCTGACGACAAACGATGATCAAGATGACC 421
Db 128 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 147

Qy ,
422 AGCTGAAGGACTTCAACTTCAACACGCCGGCTTCGACTTCTACGAAGGTGGCGTATACT 481
:: :|||::|||::|||:: :|||::|||::|||:: ||| |||
Db
148 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrLysLysProThrThr 167

RESULT 7

```

US-09-588-995A-20
; Sequence 20, Application US/09588995A
; Patent No. 6514697
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: BARNES, DEBRA A.
; APPLICANT: NELSON, RICHARD C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: METHODS FOR DETECTION
; TITLE OF INVENTION: ISOLATES AND FOR DIAG
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-5
; CURRENT APPLICATION NUMBER: US/09/588,995A
; CURRENT FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: 08/827,171
; PRIOR FILING DATE: 1997-03-27
; PRIOR APPLICATION NUMBER: 08/928,361
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: 08/700,651
; PRIOR FILING DATE: 1996-08-14
; PRIOR APPLICATION NUMBER: 08/415,751
; PRIOR FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 249

```

```

; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
US-09-588-995A-20

```

Alignment Scores:		
Pred. No.:	5.22e-07	Length:
Score:	138.50	Matches:
Percent Similarity:	43.75%	Conservative:
Best Local Similarity:	27.50%	Mismatches:
Query Match:	16.09%	Indels:
DB:	2	Gaps:
		249

US-10-049-750-13 (1-483) x US-09-588-995A-20 (1-249)

[illegible]

RESULT 9

[illegible]

RESULTS

```

US-08-700-651-5
; Sequence 5, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIB
; TITLE OF INVENTION: FOR PROPHYLAXIS
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-4 (HV)
; CURRENT APPLICATION NUMBER: US/08/7
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1721
; TYPE: PRP
; ORGANISM: Cryptosporidium parvum

```

ORGANISM: Cryptosporidium parvum
US-08-700-651-5

Alignment Scores:	
Pred. No.:	1.16e-06
Score:	138.50
Length:	1721
Percent Similarity:	43.12%
Matches:	44
Best Local Similarity:	Conservative: 25
Query Match:	Mismatches: 90
DB:	Indels: 1
	Gaps: 1
	2

US-10-049-750-13 (1-483) x US-08-700-651-5 (1-1721)

[illegible]

RESULT 9

```
US-08-928-361B-6
; Sequence 6, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1 (HV)
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1721 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-928-361B-6

Alignment Scores:
Pred. No.: 1.16e-06 Length: 1721
Score: 138.50 Matches: 44
Percent Similarity: 43.12% Conservative: 25
Best Local Similarity: 27.50% Mismatches: 90
Query Match: 16.09% Indels: 1
DB: 2 Gaps: 1

US-10-049-750-13 (1-483) x US-08-928-361B-6 (1-1721)
QY 2 TATACATAATGCAAAAAGACGATCTACTCGGTGCCGGTGTTCACCTGACCGCCAAA 61
Db 183 TyrThrLysCysValGlyValLysHisThrThrThrThrThrThrThrThrThrThr 202
QY 62 ACAAGCCCTACAAAGGAGCCATGGAAGCCCTCAAGGAAAAACCAACGATTGACCTGGAAA 121
Db 203 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 222
QY 122 ACAGCTACGTTCCCTGGACACACAGTACAAAGGTTATCCGGTGTGATGAACACCGGAAAT 181
Db 223 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 242
QY 182 ACTGCATACAAAGGTTTGGCTACGCCACCTACAAACACGACTTGAAACGGGATCAAGA 241
Db 243 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 262
QY 242 CCAACGACATCATGTGGGTGTCTACATCCCTGACGAAAGACGTCGGCTGGGCATGG 301

US-09-588-995A-6
; Sequence 6, Application US/09588995A
; Patent No. 6514697
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: BARNES, DEBRA A.
; APPLICANT: NELSON, RICHARD C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: METHODS FOR DETECTION OF CRYPTOSPORIDIUM SPECIES AND
; TITLE OF INVENTION: ISOLATES AND FOR DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-5
; CURRENT APPLICATION NUMBER: US/09/588,995A
; CURRENT FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: 08/827,171
; PRIOR FILING DATE: 1997-03-27
; PRIOR APPLICATION NUMBER: 08/928,361
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: 08/700,651
; PRIOR FILING DATE: 1996-08-14
; PRIOR APPLICATION NUMBER: 08/415,751
; PRIOR FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1721
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
US-09-588-995A-6

Alignment Scores:
Pred. No.: 1.16e-06 Length: 1721
Score: 138.50 Matches: 44
Percent Similarity: 43.12% Conservative: 25
Best Local Similarity: 27.50% Mismatches: 90
Query Match: 16.09% Indels: 1
DB: 2 Gaps: 1

US-10-049-750-13 (1-483) x US-09-588-995A-6 (1-1721)
QY 2 TATACATAATGCAAAAAGACGATCTACTCGGTGCCGGTGTTCACCTGACCGCCAAA 61
Db 183 TyrThrLysCysValGlyValLysHisThrThrThrThrThrThrThrThrThrThr 202
QY 62 ACAAGCCCTACAAAGGAGCCATGGAAGCCCTCAAGGAAAAACCAACGATTGACCTGGAAA 121
Db 203 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 222
QY 122 ACAGCTACGTTCCCTGGACACACAGTACAAAGGTTATCCGGTGTGATGAACACCGGAAAT 181
Db 223 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 242
QY 182 ACTGCATACAAAGGTTTGGCTACGCCACCTACAAACACGACTTGAAACGGGATCAAGA 241
Db 243 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 262
QY 242 CCAACGACATCATGTGGGTGTCTACATCCCTGACGAAAGACGTCGGCTGGGCATGG 301
```


THIS PAGE BLANK (USPTO)

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	126.5	14.7	288	3	US-09-216-393-341		Sequence 341, App
2	126.5	14.7	288	3	US-09-216-393-344		Sequence 344, App
3	126.5	14.7	288	4	US-10-321-856-341		Sequence 341, App
4	126.5	14.7	288	4	US-10-321-856-344		Sequence 344, App
5	126	14.6	1795	6	US-11-097-143-36210		Sequence 36210, A
6	125.5	14.6	605	3	US-09-801-368-428		Sequence 428, App
7	125.5	14.6	605	4	US-10-369-493-22016		Sequence 22016, A
8	124	14.4	292	4	US-10-425-114-59190		Sequence 59190, A
9	119.5	13.9	19723	4	US-10-084-846A-5		Sequence 5, Appli
10	119	13.8	324	4	US-10-425-114-53603		Sequence 53603, A
11	119	13.8	347	4	US-10-425-114-56484		Sequence 56484, A


```
QY 68 CTTACAGGAGCCATGGAAGCCCTCAAGGAAACCAACGATGACCTGGGAAACAGCT 127
D 159 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 178
QY 128 ACCTTCCCTGGCAACACGATACAAAGGATACCGGGTGTGATGAACACCGGAATACCTGC 187
D 179 ThrThrProThrThrThrThrThrThrThrThrThrThrThrProThrThrThrThr 198
QY 188 ATGACAGAGGTTGGGCTACCGCCACCTACAAACAGCTTGAACGGGATCAAGACCAACG 247
D 199 ThrThr-----ThrThrProThrThrThrThrThrThrThrThrThrThrProThr 215
QY 248 ACATCATGCTGGGTGTCTACATCCCTGACGACCAAGAGAGCTGCGGCTGGGATGGAACCTGG 307
D 216 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 225
QY 308 GTTACGCTTGAAGCAAGCAAGTACCTCTTTTGGTCATCCCGGAGCAAGACTACGGCA 367
D 242 ThrThrThrThrThrThrThrThrThrThrThrGluProThrThrThrThrThrThr 261
QY 428 AGGACTTCAACTTCAACAAAGCCGCTTCGACTTCTACGAAGGTGCGGTATACT 481
D 262 ThrThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrProSerThr 279

RESULT 2
US-09-216-393-344
; Sequence 344, Application US/09216393
; Patent No. US2001001447A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393
; EARLIER FILING DATE: 1998-12-18
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 344
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Toxoplasma gondii
US-09-216-393-344

Alignment Scores:
Pred. No.: 0.000556 Length: 288
Score: 126.50 Matches: 41
Percent Similarity: 43.67% Conservative: 28
Best Local Similarity: 25.95% Mismatches: 72
Query Match: 14.69% Indels: 17
DB: 3 Gaps: 2

US-10-049-750-13 (1-483) x US-09-216-393-344 (1-288)
QY 8 AAATGCCAAAGACAGCATCTACTTCGGTCCGCGCTGTTCACTACGCGCAAAACAAAG 67
D 139 ArgSerLysArgGlyLysLysThrThrThrThrThrThrThrThrThrThrThrThr 158
QY 68 CCTACAGGAGCCATGGAAGCCCTCAAGGAAACCAACGATGACCTGGGAAACAGCT 127
D 159 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 178
QY 128 ACCTTCCCTGGCAACACGATACAAAGGATACCGGGTGTGATGAACACCGGAATACCTGC 187
D 179 ThrThrProThrThrThrThrThrThrThrThrThrThrThrProThrThrThrThr 198
QY 188 ATGACAGAGGTTGGGCTACCGCCACCTCAAGGAAACCAACGATGACCTGGGAAACAGCT 247
D 199 ThrThr-----ThrThrProThrThrThrThrThrThrThrThrThrThrProThr 215
QY 248 ACCTTCCCTGGCAACACGATACAAAGGATACCGGGTGTGATGAACACCGGAATACCTGC 187
D 179 ThrThrProThrThrThrThrThrThrThrThrThrThrThrProThrThrThrThr 198
QY 188 ATGACAGAGGTTGGGCTACCGCCACCTCAACAAACAGCTTGAACGGGATCAAGACCAACG 247
D 226 -----ThrThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrPro 241
```

```
Db 199 ThrThr-----ThrThrProThrThrThrThrThrThrThrThrThrThrProThr 215
QY 248 ACATCATGCTGGGTGTCTACATCCCTGACGACCAAGAGAGCTGCGGCTGGGATGGAACCTGG 307
D 216 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 225
QY 308 GTTACGCTTGAAGCAAGCAAGTACCTCTTTTGGTCATCCCGGAGCAAGACTACGGCA 367
D 226 -----ThrThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrPro 241
QY 368 AGCGCATCAACTCATGAGCTGGGGCTCAGCGCAACAGCTGATCAAGATGAGCCAGCTGA 427
D 242 ThrThrThrThrThrThrThrThrThrThrThrGluProThrThrThrThrThrThr 261
QY 428 AGGACTTCAACTTCAACAAAGCCGCTTCGACTTCTACGAAGGTGCGGTATACT 481
D 262 ThrThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrProSerThr 279

RESULT 3
US-10-321-856-341
; Sequence 341, Application US/10321856
; Publication No. US20030194393A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
; FILE REFERENCE: TX-1-C2-1
; CURRENT APPLICATION NUMBER: US/10/321,856
; CURRENT FILING DATE: 2002-12-17
; PRIOR APPLICATION NUMBER: 09/216,393
; PRIOR FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: 08/994,825
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 366
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 341
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Toxoplasma gondii
US-10-321-856-341

Alignment Scores:
Pred. No.: 0.000556 Length: 288
Score: 126.50 Matches: 41
Percent Similarity: 43.67% Conservative: 28
Best Local Similarity: 25.95% Mismatches: 72
Query Match: 14.69% Indels: 17
DB: 4 Gaps: 2

US-10-049-750-13 (1-483) x US-10-321-856-341 (1-288)
QY 8 AAATGCCAAAGACAGCATCTACTTCGGTCCGCGCTGTTCACTACGCGCAAAACAAAG 67
D 139 ArgSerLysArgGlyLysLysThrThrThrThrThrThrThrThrThrThrThrThr 158
QY 68 CCTACAGGAGCCATGGAAGCCCTCAAGGAAACCAACGATGACCTGGGAAACAGCT 127
D 159 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 178
QY 128 ACCTTCCCTGGCAACACGATACAAAGGATACCGGGTGTGATGAACACCGGAATACCTGC 187
D 179 ThrThrProThrThrThrThrThrThrThrThrThrThrThrProThrThrThrThr 198
QY 188 ATGACAGAGGTTGGGCTACCGCCACCTCAACAAACAGCTTGAACGGGATCAAGACCAACG 247
D 199 ThrThr-----ThrThrProThrThrThrThrThrThrThrThrThrThrProThr 215
QY 248 ACATCATGCTGGGTGTCTACATCCCTGACGACCAAGAGAGCTGCGGCTGGGATGGAACCTGG 307
D 216 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 225
QY 308 GTTACGCTTGAAGCAAGCAAGTACCTCTTTTGGTCATCCCGGAGCAAGACTACGGCA 367
D 226 -----ThrThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrPro 241
```

```
QY 368 AGCGATCAACTCATGAGCTGGGCGTCACGACAACTGATCAAGATGAGCGAGCTGA 427
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 242 ThrThrThrThrThrThrThrGluProThrThrThrThrThrGluProThrThr 261
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 428 AGGACTTCAACTCAACAAAGCGCGCTTCGACTTCTACGAAGGTGCGGTACT 481
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 262 ThrThrSerThrThrThrThrThrThrThrThrThrThrThrThrProSerThr 279

RESULT 4
US-10-321-856-344
; Sequence 344, Application US/10321856
; Publication No. US20030194393A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THEREOF
; FILE REFERENCE: TX-1-C2-1
; CURRENT APPLICATION NUMBER: US/10/321,856
; CURRENT FILING DATE: 2002-12-17
; PRIOR APPLICATION NUMBER: 09/216,393
; PRIOR FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: 08/994,825
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 366
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 344
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Toxoplasma gondii
US-10-321-856-344

Alignment Scores:
Pred. No.: 0.000556 Length: 288
Score: 126.50 Matches: 41
Percent Similarity: 43.67% Conservative: 28
Best Local Similarity: 25.95% Mismatches: 72
Query Match: 14.63% Indels: 17
DB: 4 Gaps: 2

US-10-049-750-13 (1-483) x US-10-321-856-344 (1-288)
QY 8 AATATCCAAAAGACGATCTACTTCGGTGGCGGTGTTCACTACCGCCAAACAAG 67
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 139 ArgSerLysArgGlyLysThrThrThrThrThrThrThrThrThrThrThrSerThr 158
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 68 CTTCAAGAGAGCCATGGAAGCCCTCAAGGAAACCAACGATTCACCTGGGAAACAGCT 127
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 159 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 178
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 128 ACGTTCCCTGGACCAACGATCAAGGGTATCCGGTTGATGAACACCGGAAATACCTGC 187
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 179 ThrThrProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 198
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 188 ATGACAAAGTTGGGTACGGCCACTCAACAAAGCTTGAACGGGATCAAGACCAAG 247
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 199 ThrThr-----ThrThrProThrThrThrThrThrThrThrThrThrThrThr 215
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 248 ACATCATGTGGGTGTACATCCCTGACGAGAGAGAGAGCTCGCGCTGGCATGGAAC 307
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 216 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 225
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 308 GTTAGCGCTTGACCAAGCAAGTACGTCTTTTGGTCATCCCGGACGAAGACTACGGA 367
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 226 -----ThrThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrPro 241
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 368 AGCGATCAACTCATGAGCTGGGCGTCACGACAACTGATCAAGATGAGCGAGCTGA 427
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 242 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 261
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 428 AGGACTTCAACTCAACAAAGCGCGCTTCGACTTCTACGAAGGTGCGGTACT 481
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 262 ThrThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 279
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
```

RESULT 5

```
US-11-097-143-36210
; Sequence 36210, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36210
; LENGTH: 1795
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-36210

Alignment Scores:
Pred. No.: 0.00107 Length: 1795
Score: 126.00 Matches: 50
Percent Similarity: 47.02% Conservative: 29
Best Local Similarity: 29.76% Mismatches: 73
Query Match: 14.63% Indels: 16
DB: 6 Gaps: 5

US-10-049-750-13 (1-483) x US-11-097-143-36210 (1-1795)
QY 5 ACTAATCCCAAAAAGACGATCTACTTCGGTGGCGGTGTTCACTACCGCCAAACA 64
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 493 ThrLysSerThrProLysIleSerSerThrThrGluGlnHisSerThrThrAlaLysThr 512
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 65 AAGCCTCAAGAAAGCCATGGAAGCCCTCAAGGAAACCAACGATTCACCTGGAAACA 124
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 513 ThrThrThrLysArgProThrThrValThrGluLysThrSerSerAlaThrGluLysPro 532
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 125 GCTACGTTCCCTCGCAACCAACGATCA-----AGGGTATCCGGGTGATGAACACC--- 175
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 533 ArgThrThrValThrThrThrThrGlnLysArgSerThrThrHisAsnThrSer 552
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 176 -----CGAATACCTGCATCAACAGGTTTGGGCTACGCCACCTACCAACAGACTTGA 229
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 553 ProAspThrLysThrThrIleArgSerThrThrLeuSerProLysThrThrThrPro 572
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 230 ACGGATCAAGACCA-----ACGACATCATGCTGGGTGCTACATCCCTGCACGAAGAAGCG 286
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 573 SerThrThrProSerThrThrThrProSerThrThrThrProSerThrThrThrPro 592
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 287 TCGGCTGGCATGGAAGTACGCTTACGCTTGAAGCAAGGACGATCGCTCTTTGGTCA 346
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 593 SerThrThrThrProSerThrThrThrProSerThrThrThrValSerThr 612
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 347 TCCCGGACGAAGACTACGGCAAGCGCATCAACCTCATGAGCTGGGGCGTCACGCAACG 406
   ::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
```

Db 613 HisArgProArgThrThrSerGlnLysThrThr-----AlaSerThrThrThr 629
QY 407 TGATCAAGATGAGCCAGCTGAAGGACT-----TCAACTTCAACA 445
Db 630 LysLysThrThrThrSerProLysThrThrLysThrThrAspIleProThrSerThrThr 649
QY 446 AGCGCGCTTCGACTTCTACGAG 469
Db 650 SerLysLeuSerThrThrThrGln 657

RESULT 6
US-09-801-368-428
; Sequence 428, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 428
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-428

Alignment Scores:
Pred. No.: 0.000878 Length: 605
Score: 125.50 Matches: 51
Percent Similarity: 45.96% Conservative: 23
Best Local Similarity: 31.68% Mismatches: 48
Query Match: 14.58% Indels: 39
DB: 3 Gaps: 7

US-10-049-750-13 (1-483) x US-09-801-368-428 (1-605)

QY 5 ACTAAATGCCMAAAGAGCATCTACTTCGGTCCGGCTGGTTCAGTCAAGCCCAAAACA 64
Db 181 ThrThrThrSerThrLysLeuSerThrSerIleProThr---SerThrThrSerThr 199
QY 65 AAGCCTACAGGAGCCATGGAAGCCCTCAAGGAAACCAACGATTGACCTGGAACA 124
Db 200 SerThrThrThrSerThrSerSerThrThrThrValSerValThrSerSerThr 219
QY 125 GCTACGTTCCCTCGACCAACAGTACAGGATATCCGGTGTGATGAACACCCGGAATACC 184
Db 220 SerThr-----ThrThrSerThrThrSerThrLeuIleSerThr----- 233
QY 185 TGCATGACAGGTTGGGCTACGGCCACTACAGCAACAGCATTCGAACGGATCAAGACCA 244
Db 234 -----SerThrSerSerSerSerSerSerThrPro 243
QY 245 ACGACA-----TCATGCTGGGTGTCTACATCCCTGACGAAGAGACGCTCGGCTGGCA 298
Db 244 ThrThrThrSerSerAlaProIleSerThrThrThrSerThrSerThrThr----- 261
QY 299 TGGNACTGGGTACGCTTCGAGCCAGGCAAGTACGTCCTTTGGTTCATCCCGACCAAG 358
Db 244 ThrThrThrSerSerAlaProIleSerThrThrThrSerThrSerThrThr----- 261

QY 299 TGGNACTGGGTACGCTTCGAGCCAGGCAAGTACGTCCTTTGGTTCATCCCGACCAAG 358
Db 262 -----SerThrSerThrThrThrSerProThrSerSerSerAlaPro 274
QY 359 ACTACGGCAAGCGCATCAACCTCATGAGCTGGGGCGCTCAGCAGCAACGTCATCAAGATGA 418
Db 275 ThrSerSerSerAsnThrThr-----ProThrSerThrThr----- 286
QY 419 GCCAGCTGAAGGACTTCAACTTCAACAGCGCGCTTCGACTTCTACGAAGGTGCGGTAT 478
Db 287 -----PheThrThrThrSerProSerThrAlaProSerSerThrThrValThrTyr 303
QY 479 ACT 481
Db 304 Thr 304

RESULT 7
US-10-369-493-22016
; Sequence 22016, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22016
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22016

Alignment Scores:
Pred. No.: 0.000878 Length: 605
Score: 125.50 Matches: 51
Percent Similarity: 45.96% Conservative: 23
Best Local Similarity: 31.68% Mismatches: 48
Query Match: 14.58% Indels: 39
DB: 4 Gaps: 7

US-10-049-750-13 (1-483) x US-10-369-493-22016 (1-605)

QY 5 ACTAAATGCCMAAAGAGCATCTACTTCGGTCCGGCTGGTTCAGTCAAGCCCAAAACA 64
Db 181 ThrThrThrSerThrLysLeuSerThrSerIleProThr---SerThrThrSerThr 199
QY 65 AAGCCTACAGGAGCCATGGAAGCCCTCAAGGAAACCAACGATTGACCTGGAACA 124
Db 200 SerThrThrThrSerThrSerSerThrThrThrValSerValThrSerSerThr 219
QY 125 GCTACGTTCCCTCGACCAACAGTACAGGATATCCGGTGTGATGAACACCCGGAATACC 184
Db 220 SerThr-----ThrThrSerThrThrSerSerThrLeuIleSerThr----- 233
QY 185 TGCATGACAGGTTGGGCTACGGCCACTACAGCAACAGCATTCGAACGGATCAAGACCA 244
Db 234 -----SerThrSerSerSerSerSerSerThrPro 243
QY 245 ACGACA-----TCATGCTGGGTGTCTACATCCCTGACGAAGAGACGCTCGGCTGGCA 298
Db 244 ThrThrThrSerSerAlaProIleSerThrThrThrSerThrSerThrThr----- 261
QY 299 TGGNACTGGGTACGCTTCGAGCCAGGCAAGTACGTCCTTTGGTTCATCCCGACCAAG 358
Db 244 ThrThrThrSerSerAlaProIleSerThrThrThrSerThrSerThrThr----- 261

Db 262 -----SerThrSerThrThrThrSerProThrSerSerSerAlaPro 274
QY 359 ACTACGGCAAGCCGATCAACCTCATGAGCTGGGGCGTCAGGACAAAGTGATCAAGATGA 418
Db 275 ThrSerSerSerAsnThrThr-----ProThrSerThrThr----- 286
QY 419 GCCAGCTGAAGGACTTCAACTTCAACAAGCCGCGTTCGACTTCTACGAAGTGCGGAT 478
Db 287 -----PheThrThrThrSerProSerThrAlaProSerSerThrThrValThrTyr 303
QY 479 ACT 481
Db 304 Thr 304
RESULT 8
US-10-425-114-59190
; Sequence 59190, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: Plants and Uses Thereof for Plant Improvement
; CURRENT APPLICATION NUMBER: US/10/425,114
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 59190
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700377772_FLI.pap
US-10-425-114-59190

Alignment Scores:
Pred. No.: 0.00101 Length: 292
Score: 124.00 Matches: 54
Percent Similarity: 42.5% Conservative: 26
Best Local Similarity: 28.7% Mismatches: 62
Query Match: 14.4% Indels: 46
DB: 4 Gaps: 11

US-10-049-750-13 (1-483) x US-10-425-114-59190 (1-292)

QY 14 CAAATAAGACGATCTACTTTCGGTGGCGGTGT-----TCAGTACGCCCAAAACA 64
Db 21 GlnGluArgThrSerThrThrGlnProSerSerThrProThrArgCysAlaArgAla 40
QY 65 AGCCCTACAAAG-----AGCCATCGAAGCCC-----TCAGGAAACCAACGNTTG 112
Db 41 ArgProSerArgProLeuAspProThrArgGlyCysTyrArgArgProArgGlnArgPro 60
QY 113 ACCTGGAACACAGTACGTCCTCCCTGGACACAGTACAAAGGGTATCCGGG-----TTG 166
Db 61 SerTppProProSer-----TrpTrpArgAlaCysArgAlaProGlyArgArgLeu 77
QY 167 ATGAACACCCGAATACCTGCATGACAAAGTTTGGGCTA----- 205
Db 78 ProArgArgArgThrThrArgArgThrGlyThrThrProLeuAlaAlaMet 97
QY 206 -----CGCCCACTACACACGACTTGAACGGGA 235
Db 98 GlyGlySerSerThrProThrProThrProThrProGlyThrThrProThrGlyProPro 117
QY 236 TCAAGACCAACGACATCATGCTGGGTGTCTACATCCTGACGAAGACGTCGGCTGG 295
Db 118 AlaArgProSerThrSer-----ValThrThrSerPheSerArgArgThrThrThrArg 135

QY 296 GCATGG-----AACTGGGTTCAGCCCTTGAGCCAAGCA 328
Db 136 ArgTrpCysArgProProThrProArgHisSerGlySerAlaThrPro-----Ala 152
QY 329 AGTAGCTCCTTTTGGTTCATCCCGGACGAGAGACTACGGCAAGCCGATCAACCTCATGAGCT 388
Db 153 ThrThr-----TrpArgArgArgProSerThrAlaAlaAlaAlaGlyAlaAla 170
QY 389 GGGCGGTTCAGCAGCAACGTGATCAAGATGAGCCAGC-----TGAAGGACTTCAACTTCAACA 445
Db 171 AlaSerSerArgThrAlaProSerArgTrpArgSerProProArgAlaProThrThrSer 190
QY 446 AGCCGCGCTTCGACTTCTACGAAG 469
Db 191 SerProArgProThrGlyAlaArg 198

RESULT 9

US-10-084-846A-5
; Sequence 5, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLNIEWEG, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 5
; LENGTH: 19723
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: att, Start position: nucleotide 3.

US-10-084-846A-5

Alignment Scores:
Pred. No.: 0.0102 Length: 19723
Score: 119.50 Matches: 54
Percent Similarity: 28.1% Conservative: 16
Best Local Similarity: 21.6% Mismatches: 62
Query Match: 13.8% Indels: 117
DB: 4 Gaps: 9

US-10-049-750-13 (1-483) x US-10-084-846A-5 (1-19723)

QY 11 TGCCAAAAGACGATCTACTTCGGTGGCGGTGTCTCAGTCACGCCCAAAACAAGCT 70
Db 9192 CysArgProAlaTrpThrThrCysProAlaGlySerArgThrSerThrArgSerPro 9211
QY 71 ACAAGGAAG----- 79
Db 9212 ThrArgSerGlySerArgCysGlyProThrProSerArgSerProSerArgGlyArg 9231
QY 80 -----CCATGG----- 85
Db 9232 SerThrThrSerSerThrTrpArgArgProProProTrpThrIleTrpArgCysArg 9251
QY 86 -----AAGCCCTCAAGGAAACCCAA---CGATTGACCTGGAAA 121
Db 9252 TrpArgArgCysGlySerAlaArgProGlyProArgThrArgCysGlyTrpArgTrpArg 9271
QY 122 ACAGCTACGTTCCCTCG-----ACA 142
Db 9272 ThrAlaProAlaSerTrpSerProProArgSerThrGlyIleProProSerThr 9291

```

QY 143 ACCAGTACAGGGTATCCGGTTGATGAACACCCGGATACC-----TGCATGACA 193
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9292 ArgSerProArgAlaThrGlyAlaThrSerThrArgSerAlaArgValArgCysThrThr 9311
QY 194 AGG-----TTTGGGTACGGCCACCT----- 214
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9312 ArgProAsnGlySerProArgArgProProProThrArgAlaArgTTPAlaArgThrPro 9331
QY 215 -----ACAACAACGACTTGAACCGGGATCAAGACCAACGACATCATCTGCG 259
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9332 AlaLeuProAlaCysSerThrProThrAspArgGlyCysAlaAlaThrThrGlyGlyTip 9351
QY 260 GTGTCTACATCCCTGACGAG----- 280
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9352 CysArgProSerSerThrArgArgSerProAspCysArgSerProSerThrAlaAlaAla 9371
QY 281 -----AAGACGTCCGGCTGGGCATCGAATCTGGTTACG 313
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9372 ProArgProGlyArgCysAlaThrSerArgThrArgCysGlyAspTPProTPThrArg 9391
QY 314 CTTGAGCCCAAGGCAAGTACGTCTTT----- 340
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9392 ProSerArgAlaArgSerThrSerAlaArgProValArgProSerValArgSerArgArg 9411
QY 341 -----TGTCATCCCGGACGAGACTAGCGCAACCGCATCAACC 379
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9412 ArgSerProHisSerArgGluTPSerCysGlyProSerSerAlaArgArgArgThr 9431
QY 380 TCATGAGCTGGGGCGTCAGCGACAACG 406
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9432 SerProAlaGlyAlaAlaProThrSer 9440

```

RESULT 10

```

US-10-425-114-53603
; Sequence 53603, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 53603
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17202D01_FLI.pep
US-10-425-114-53603

```

```

Alignment Scores:
Pred. No.: 0.00345 Length: 324
Score: 119.00 Matches: 48
Percent Similarity: 42.47% Conservative: 14
Best Local Similarity: 32.88% Mismatches: 44
Query Match: 13.82% Indels: 40
DB: 4 Gaps: 9

```

```

US-10-049-750-13 (1-483) x US-10-425-114-53603 (1-324)

```

```

QY 75 GGAAGCCATGAAGCCCTCAAGAAAACCCCAAGATTGACCTGGAAACAG----- 125
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155 GlyLeuHisGlyGlyArgArgAlaLeu-----ProGlyArgHisArgArgAla 170
QY 126 -----CTAGTTCCTCCCTGGACACCAAGGTATATCCG 161

```

```

Db 171 AlaAlaArgGlnArgValProLeuLeuArgValProGlyGlnProAlaArgHisGlnPro 190
QY 162 GGTGTGATGAACACCCGGAATACCTGCATGACAAAGTTTGGGTACGCGCCAC----- 212
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
191 Gly-----LeuArgHisValProAla-----GlyHisHisArgGln 202
QY 213 ---CTACAACAAGACTTGAACCGGATCAAGACCAACGACATCATCTGGTGTCTACAT 269
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
203 GlyGlnArgGlnArgProGlnLeuGlnProValArgArgHisGlyGly----- 219
QY 270 CCCTGACGAAGAAGACGTGGCGCTGGCATGGTTCAGCTTGGCTTGAAGCCCAAGSCAA 329
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
220 -----ArgArgArgArgArgAlaArgGluGlyArgAlaGlnArgGlnGlyArgArg 237
QY 330 GTACGTCTTTTGGTCATCCCGGACGAAGACTACCG-----CAAGCCGAT 374
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
238 ValArgGluArgLeuAlaLeuGlyArgArgLeuArgGlyGlnArgArgGlnArgAlaGlu 257
QY 375 ---CAACCT-----CATGAGCTGGGGCGTCACGACCAACGTCATCAAGATGAGCCAGCT 425
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
258 IleGlnProGlyThrHisArgProArgArgProArgHisAlaGlnGluAspArgProAla 277
QY 426 GAAGGACTTCAACTTCAA 443
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
278 GlyAspLeuArgLeuArg 283

RESULT 11
US-10-425-114-56484
; Sequence 56484, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 56484
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73213B10_FLI.pep
US-10-425-114-56484

```

```

Alignment Scores:
Pred. No.: 0.00352 Length: 347
Score: 119.00 Matches: 48
Percent Similarity: 42.47% Conservative: 14
Best Local Similarity: 32.88% Mismatches: 44
Query Match: 13.82% Indels: 40
DB: 4 Gaps: 9

```

```

US-10-049-750-13 (1-483) x US-10-425-114-56484 (1-347)

```

```

QY 75 GGAAGCCATGAAGCCCTCAAGAAAACCCCAAGATTGACCTGGAAACAG----- 125
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
178 GlyLeuHisGlyGlyArgArgAlaLeu-----ProGlyArgHisArgArgAla 193
QY 126 -----CTAGTTCCTCCCTGGACACCAAGGTATATCCG 161
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
194 AlaAlaArgGlnArgValProLeuLeuArgValProGlyGlnProAlaArgHisGlnPro 213
QY 162 GGTGTGATGAACACCCGGAATACCTGCATGACAAAGTTTGGGTACGCGCCAC----- 212
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
214 Gly-----LeuArgHisValProAla-----GlyHisHisArgGln 225

```

```
Qy 213 ---CTACAAACGACCTTGAACGGGATCAAGACCAACGACATCATGCTGGGTCTCAT 269
Db 226 GlyGlnArgGlnArgProGlnLeuGlnGlnProValArgArgHisGlyGly 242
Qy 270 CCCTGACGAACGACCTGGCCCTGGGCGATGGAACCTGGTTACGCTTGAGCCAAAGCAA 329
Db 243 -----ArgArgArgArgArgAlaArgAlaArgGluGlyArgArgAlaGlnArgGlnGlyArg 260
Qy 330 GTACGTCCTTTTGGTCTATCCCGAGCAAGACTACGG-----CAAGCCGAT 374
Db 261 ValArgGluArgLeuAlaLeuGlyArgArgLeuArgGlyGlnArgArgGlnArgAlaGlu 280
Qy 375 ---CAACCT-----CATGACCTGGGCGGTGAGCGACACCTGATCAAGATGAGCCAGCT 425
Db 281 IlegInProGlyThrHisArgProArgArgProArgHisAlaGlnGluAspArgProAla 300
Qy 426 GAAGGACTTCAACTTCAA 443
Db 301 GlyAspLeuArgLeuArg 306

RESULT 12
US-10-425-114-59602
; Sequence 59602, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 59602
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3594-027-AL_FLI.pep
US-10-425-114-59602

Alignment Scores:
Pred. No.: 0.00356 Length: 361
Score: 119.00 Matches: 48
Percent Similarity: 42.47% Conservative: 14
Best Local Similarity: 32.88% Mismatches: 44
Query Match: 13.82% Indels: 40
DB: 4 Gaps: 9

US-10-049-750-13 (1-483) x US-10-425-114-59602 (1-361)
Qy 75 GGAAGCCATGGAAGCCCTCAAGGAAACCCCAACGATTGACCTGGAAGAACAG----- 125
Db 192 GlyLeuHisGlyGlyArgArgAlaLeu-----ProGlyArgHisArgAla 207
Qy 126 -----CTACGTTCCCTGGACACACCTGACCAAGGTTGGGCTACGGCCAC----- 212
Db 208 AlaAlaArgGlnArgValProLeuLeuArgValProGlyGlnProAlaArgHisGlnPro 227
Qy 162 GGTTCATGACACCCGGAATACCTGCATGCAAGGTTGGGCTACGGCCAC----- 212
Db 228 Gly-----LeuArgHisValProAla-----GlyHisArgGln 239
Qy 213 ---CTACAAACGACCTTGAACGGGATCAAGACCAACGACATCATGCTGGGTCTCAT 269
Db 240 GlyGlnArgGlnArgProGlnLeuGlnGlnProValArgArgHisGlyGly----- 256
Qy 270 CCCTGACGAAGAAGACGTCGGCCCTGGGCGATGGAACCTGGAAGTTCAGCCTTGAGCCAAAGCAA 329
```

```
Db 257 -----ArgArgArgArgArgAlaArgGluGlyArgArgAlaGlnArgGlnGlyArgArg 274
Qy 330 GTACGTCCTTTTGGTCTATCCCGAGCAAGACTACGG-----CAAGCCGAT 374
Db 275 ValArgGluArgLeuAlaLeuGlyArgArgLeuArgGlyGlnArgArgGlnArgAlaGlu 294
Qy 375 ---CAACCT-----CATGACCTGGGCGGTGAGCGACACCTGATCAAGATGAGCCAGCT 425
Db 295 IlegInProGlyThrHisArgProArgArgProArgHisAlaGlnGluAspArgProAla 314
Qy 426 GAAGGACTTCAACTTCAA 443
Db 315 GlyAspLeuArgLeuArg 320

RESULT 13
US-10-425-114-68177
; Sequence 68177, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 68177
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3594-049-B7_FLI.pep
US-10-425-114-68177

Alignment Scores:
Pred. No.: 0.00356 Length: 362
Score: 119.00 Matches: 48
Percent Similarity: 42.47% Conservative: 14
Best Local Similarity: 32.88% Mismatches: 44
Query Match: 13.82% Indels: 40
DB: 4 Gaps: 9

US-10-049-750-13 (1-483) x US-10-425-114-68177 (1-362)
Qy 75 GGAAGCCATGGAAGCCCTCAAGGAAACCCCAACGATTGACCTGGAAGAACAG----- 125
Db 193 GlyLeuHisGlyGlyArgArgAlaLeu-----ProGlyArgHisArgAla 208
Qy 126 -----CTACGTTCCCTGGACACACCTGACCAAGGTTGGGCTACGGCCAC----- 212
Db 209 AlaAlaArgGlnArgValProLeuLeuArgValProGlyGlnProAlaArgHisGlnPro 228
Qy 162 GGTTCATGACACCCGGAATACCTGCATGCAAGGTTGGGCTACGGCCAC----- 212
Db 229 Gly-----LeuArgHisValProAla-----GlyHisArgGln 240
Qy 213 ---CTACAAACGACCTTGAACGGGATCAAGACCAACGACATCATGCTGGGTCTCAT 269
Db 241 GlyGlnArgGlnArgProGlnLeuGlnGlnProValArgArgHisGlyGly----- 257
Qy 270 CCCTGACGAAGAAGACGTCGGCCCTGGGCGATGGAACCTGGAAGTTCAGCCTTGAGCCAAAGCAA 329
Db 258 -----ArgArgArgArgArgAlaArgGluGlyArgArgAlaGlnArgGlnGlyArgArg 275
Qy 330 GTACGTCCTTTTGGTCTATCCCGAGCAAGACTACGG-----CAAGCCGAT 374
Db 276 ValArgGluArgLeuAlaLeuGlyArgArgLeuArgGlyGlnArgArgGlnArgAlaGlu 295
```

```
QY 375 ---CAACCT-----CATGAGTCGGCGCTCAGCGCAACAGTGTATCAAGATGAGCCAGCT 425
Db 296 IleGlnProGlyThrHisArgProArgArgProArgHisAlaGlnGluAspArgProAla 315
QY 426 GAAGGACTCAACTTCAA 443
Db 316 GlyAspLeuArgLeuArg 321

RESULT 14
US-10-425-114-61578
; Sequence 61578, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 61578
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3069-035-G11_FLI.pep
US-10-425-114-61578

Alignment Scores:
Pred. No.: 0.00497 Length: 333
Score: 117.50 Matches: 48
Percent Similarity: 42.77% Conservative: 20
Best Local Similarity: 30.19% Mismatches: 58
Query Match: 13.65% Indels: 33
DB: 4 Gaps: 6

US-10-049-750-13 (1-483) x US-10-425-114-61578 (1-333)
QY 23 CGATCTACTTCGGTGGCGGTCTCACTGACGCCCAAAACAAGCCT-----ACA 73
Db 107 ArgGlyThrSerThrGlySerArgThrGlySerArgThrGlySerArgThrGlySerArgThr 126
QY 74 AGNAGCCATGGAAGCCCTCAAGGAAACCCACGATTGACCTGGGAAACAGTACGTTTC 133
Db 127 ArgAlaThrArgProSerArgSerSerThrArgSerProAlaArgArgSerThr 146
QY 134 CCCTGGACACCACTACAGGGTATCCGGGTGTGATGAACACCCGGAATACCTGCATGACA 193
Db 147 ProArgGlyAlaThrProSerSerSerCysProThrThrArgThr----- 163
QY 194 AGGTTGGGTACGGCCACCTTACAACACGACTTGAACGGGATCAAGCAACCAACGACATCA 253
Db 164 -----ArgThrSerProCysSerGlyProArgAlaAlaThrSer 176
QY 254 TGCTGGGTGTCTACATCCCTGACGAAGACGTCGGCGCTGGGGATCGAACTGGTTACG 313
Db 177 AlaThrAlaSerAlaAla-----SerAlaTrpSerThr----- 187
QY 314 CCTTGAGCCAAAGCAAGTACGTCCTTTTGGTTCATCCCGGACGAGACTACGCAAGC--- 370
Db 188 -----ThrSerThrSerThrSerThrProSerThrAlaThrArgLeuThrAlaAlaSer 205
QY 371 -----CGATCAACCTCATGAGCTGGGCGCTCAGCGACCAACGATGATCAAGATGA 418
Db 206 MethrValProArgSerCysSerGlySerGlyAlaAlaThr----- 221
QY 419 GCCAGCTGAAGGACTTCAACTTCAACAGCGCGCTTCGACTTCTACGAAGGTGCCG 475
```

```
Db 222 SerAlaGlyArgSerSerLeuGlyGluArgThrLysGluThrThrSerTrpTrpPro 240

RESULT 15
US-10-425-114-53657
; Sequence 53657, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 53657
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3150-064-E8_FLI.pep
US-10-425-114-53657

Alignment Scores:
Pred. No.: 0.00518 Length: 383
Score: 117.50 Matches: 48
Percent Similarity: 42.77% Conservative: 20
Best Local Similarity: 30.19% Mismatches: 58
Query Match: 13.65% Indels: 33
DB: 4 Gaps: 6

US-10-049-750-13 (1-483) x US-10-425-114-53657 (1-383)
QY 23 CGATCTACTTCGGTGGCGGTCTCACTGACGCCCAAAACAAGCCT-----ACA 73
Db 157 ArgGlyThrSerThrGlySerArgThrGlySerArgThrGlySerArgThrGlySerGlyThr 176
QY 74 AGNAGCCATGGAAGCCCTCAAGGAAACCCACGATTGACCTGGGAAACAGTACGTTTC 133
Db 177 ArgAlaThrArgProSerArgSerSerThrArgSerProAlaArgArgSerThr 196
QY 134 CCCTGGACACCACTACAGGGTATCCGGGTGTGATGAACACCCGGAATACCTGCATGACA 193
Db 197 ProArgGlyAlaThrProSerSerSerCysProThrThrArgThr----- 213
QY 194 AGGTTGGGTACGGCCACCTTACAACACGACTTGAACGGGATCAAGCAACCAACGACATCA 253
Db 214 -----ArgThrSerProCysSerGlyProArgAlaAlaThrSer 226
QY 254 TGCTGGGTGTCTACATCCCTGACGAAGACGTCGGCGCTGGGCATCGAAGTGGTTACG 313
Db 227 AlaThrAlaSerAlaAla-----SerAlaTrpSerThr----- 237
QY 314 CCTTGAGCCAAAGCAAGTACGTCCTTTTGGTTCATCCCGGACGAGACTACGCAAGC--- 370
Db 238 -----ThrSerThrSerThrSerThrProSerThrAlaThrArgLeuThrAlaAlaSer 255
QY 371 -----CGATCAACCTCATGAGCTGGGCGCTCAGCGACCAACGATGATCAAGATGA 418
Db 256 MethrValProArgSerCysSerGlySerGlyAlaAlaThr----- 271
QY 419 GCCAGCTGAAGGACTTCAACTTCAACAGCGCGCTTCGACTTCTACGAAGGTGCCG 475
Db 272 SerAlaGlyArgSerSerLeuGlyGluArgThrLysGluThrThrSerTrpTrpPro 290

Search completed: December 23, 2005, 23:20:16
Job time : 99.1442 secs
```

THIS PAGE BLANK (USE)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 23, 2005, 22:44:48 ; Search time 3.9562 Seconds
(without alignments)
1741.646 Million cell updates/sec

Title: US-10-049-750-13

Perfect score: 861

Sequence: 1 gtatacaaatgccaaaaa.....acgaaggtgcgtatactaa 483

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 54001 seqs, 7132810 residues

Total number of hits satisfying chosen parameters: 108002

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US10049750/runat_23122005_113555_6421/app_query.fasta.1.1742
-DB=Published Applications AA New -QFMT=fastan -SUFTX=n2p.rapbn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10049750 @CGN 1.1.1 @runat_23122005_113555_6421
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA New:

1: /cgn2_6/ptodata/2/pubpaa/US08 NEW PUB.pcp.*
2: /cgn2_6/ptodata/2/pubpaa/US06 NEW PUB.pcp.*
3: /cgn2_6/ptodata/2/pubpaa/US07 NEW PUB.pcp.*
4: /cgn2_6/ptodata/2/pubpaa/PTCT NEW PUB.pcp.*
5: /cgn2_6/ptodata/2/pubpaa/US09 NEW PUB.pcp.*
6: /cgn2_6/ptodata/2/pubpaa/US10 NEW PUB.pcp.*
7: /cgn2_6/ptodata/2/pubpaa/US11 NEW PUB.pcp.*
8: /cgn2_6/ptodata/2/pubpaa/US60 NEW PUB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	95	11.0	445	US-10-793-626-1294	Sequence 1294, Ap
2	94.5	11.0	1236	US-10-873-528-109	Sequence 109, App
3	92.5	10.7	5179	US-11-108-172-1068	Sequence 1068, App
4	87.5	10.2	1076	US-10-131-826A-219	Sequence 219, App
5	85.5	9.9	761	US-10-485-517-252	Sequence 252, App
6	79.5	9.2	596	US-11-102-240-100	Sequence 100, App
7	78	9.1	750	US-11-070-627-4	Sequence 4, Appli
8	76.5	8.9	957	US-11-108-172-1065	Sequence 1065, Ap
9	76	8.8	1532	US-10-821-234-914	Sequence 914, App
10	75.5	8.8	616	US-10-858-730-21	Sequence 21, Appli

11	75	8.7	537	6	US-10-641-678-62	Sequence 62, Appli
12	74.5	8.7	605	7	US-11-094-586-4	Sequence 4, Appli
13	74.5	8.7	672	7	US-11-000-463-455	Sequence 455, App
c 14	74.5	8.6	993	7	US-11-137-465-36	Sequence 36, Appli
15	74	8.6	431	6	US-10-821-234-1065	Sequence 1065, Ap
16	74	8.6	540	6	US-10-641-678-60	Sequence 60, Appli
17	73.5	8.5	168	7	US-11-044-111-27	Sequence 27, Appli
18	73.5	8.5	252	6	US-10-793-626-1948	Sequence 1948, Ap
19	73.5	8.5	276	6	US-10-467-657-4172	Sequence 4172, Ap
20	73.5	8.5	276	6	US-10-467-657-7078	Sequence 7078, Ap
21	73	8.5	756	7	US-11-074-176-202	Sequence 202, App
22	72.5	8.4	1464	7	US-11-000-463-243	Sequence 243, App
23	72.5	8.4	1464	7	US-11-186-284-28	Sequence 28, Appli
24	72.5	8.4	1467	6	US-10-821-234-1096	Sequence 1096, Ap
c 25	72.5	8.4	1798	6	US-10-995-561-1033	Sequence 1033, Ap
c 26	72.5	8.4	1798	6	US-10-995-561-1034	Sequence 1034, Ap
27	71.5	8.3	282	7	US-11-087-177-9	Sequence 9, Appli
28	71	8.2	305	7	US-11-080-091-2	Sequence 2, Appli
29	71	8.2	305	7	US-11-087-177-7	Sequence 7, Appli
30	71	8.2	390	6	US-10-485-517-235	Sequence 235, App
31	70.5	8.2	750	7	US-11-070-627-3	Sequence 3, Appli
32	70.5	8.2	1366	6	US-10-821-234-1431	Sequence 1431, Ap
33	70.5	8.2	1366	7	US-11-186-284-31	Sequence 31, Appli
34	70	8.1	52	6	US-10-467-657-4144	Sequence 4144, Ap
c 35	70	8.1	406	6	US-10-878-556A-41	Sequence 41, Appli
c 36	70	8.1	673	7	US-11-102-240-16	Sequence 16, Appli
c 37	70	8.1	744	7	US-11-186-284-37	Sequence 37, Appli
c 38	70	8.1	744	7	US-11-186-284-39	Sequence 39, Appli
39	70	8.1	1388	6	US-10-821-234-1143	Sequence 1143, Ap
40	69.5	8.1	391	6	US-10-979-821-12	Sequence 12, Appli
c 41	69.5	8.0	638	6	US-10-485-517-260	Sequence 260, App
42	69.5	8.1	750	7	US-11-070-627-2	Sequence 2, Appli
43	69.5	8.1	786	7	US-11-070-627-9	Sequence 9, Appli
c 44	69.5	8.0	1172	7	US-11-186-284-203	Sequence 203, App
c 45	69	8.0	277	7	US-11-132-285-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-10-793-626-1294
; Sequence 1294, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:

; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US

; CURRENT APPLICATION NUMBER: US/10/793,626

; CURRENT FILING DATE: 2004-03-04

; PRIOR APPLICATION NUMBER: 60/164,258

; PRIOR FILING DATE: 1999-11-09

; NUMBER OF SEQ ID NOS: 4472

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1294

; LENGTH: 445

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic

; OTHER INFORMATION: amino acid sequence

US-10-793-626-1294

Alignment Scores:

Pred. No.: 0.0472 Length: 445
Score: 95.00 Matches: 54
Percent Similarity: 43.82% Conservative: 24
Best Local Similarity: 30.34% Mismatches: 54
Query Match: 11.03% Indels: 46
DB: 6 Gaps: 10

US-10-049-750-13 (1-483) x US-10-793-626-1294 (1-445)

QY 5 ACTAATGCCAAAAGAGGATCTACTTGGTGGCGGTGTTCTACTGACCCCAAAACA 64


```
QY 326 GCAAGTAGCTCCTTTGGTTCATCCCGGACGAAGACTACGGCA-----AGCCGATCA 376
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
723 AlaserThrGlnThrGlyLeuProAlaThrLeuThrThrAlaAspLeuGlyGluGlySer 742
QY 377 ACCTCATGAGCTGGGGCGTCAGCGACAAACGTCATGATCAAGATGAGCGACGCTGAAGGACTTCA 436
Db |||::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
743 ThrThrPheProSerSerSerGlySerThr-----GlyThrThr 755

QY 437 ACTTCAACAGCCGCTTCGACTTCT 463
Db |||::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
756 LeuSerProAlaArgSerThrThrSer 764

RESULT 9
US-10-821-234-914
; Sequence 914, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; FILE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR FILING DATE: 2004-04-07
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 914
; LENGTH: 1532
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-914

Alignment Scores:
Pred. No.: 4.7 Length: 1532
Score: 76.00 Matches: 35
Percent Similarity: 42.11% Conservative: 13
Best Local Similarity: 30.70% Mismatches: 44
Query Match: 8.83% Indels: 22
DB: 6 Gaps: 7

US-10-049-750-13 (1-483) x US-10-821-234-914 (1-1532)
QY 60 AAACAAGCCTTACAAGGAAGCCATCGAAGCCCTCAAGGAAACCCACGATTGACCTGGA 119
Db |||::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
590 LysLeuMetMetGluGlnGluAsnGlyAsnLeuArgGlySerProGly-----ProLys 607
QY 120 AAACAGTACGTTCCCTCGA---CAACCAAGTCAAGGGTATCGGGTTGATGAACACCC 176
Db :: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
608 GlyAspMetGlySerProGlyProLysGlyAspArgGlyPheProGly-----ThrPro 625
QY 177 GGAATACCTGCATGACAAAGTTGGGCTAC-----GCCACCTTACACACGA 224
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
626 GlyIleProGly-----ProLeuGlyHisProGlyProGlnGlyProLysGlyGlnLys 643
QY 225 -----CTTGAACGGGATCAAGACCAACGACATCATGCTGGGTGT 263
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
644 GlySerValGlyAspProGlyMetGluGlyProMetGlyGlnArgGlyArgGluGlyPro 663
QY 264 CTACATCCCTGAGAAAGACGTCGGCCTGGGATGGA-----ACTGGGTGA 311
Db :: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
664 MetGlyProArgGlyGluAlaGlyProProGlySerGlyGluLysGlyGluArgGlyAla 683
QY 312 CGCCTTGAGCCAGGCAAGTACGTCCTTTTGGTCATCCCGGA 353
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
684 AlaGlyGluProGlyProHisGlyProProGlyValProGly 697

RESULT 10
US-10-858-730-21
; Sequence 21, Application US/10858730
; Publication No. US20050255568A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.
; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgey, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; FILE OF INVENTION: PRODUCTION
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858,730
; CURRENT FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 616
; TYPE: PRT
; ORGANISM: Thermobifida fusca
US-10-858-730-21

Alignment Scores:
Pred. No.: 4.09 Length: 616
Score: 75.50 Matches: 40
Percent Similarity: 34.72% Conservative: 10
Best Local Similarity: 27.78% Mismatches: 42
Query Match: 8.77% Indels: 52
DB: 6 Gaps: 7

US-10-049-750-13 (1-483) x US-10-858-730-21 (1-616)
QY 14 CAAAAAGACGATCTACTTCGTCGCGCTGTTCACCTGACCGCCAAACAAAGCCTTACA 73
Db |||::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
71 GlnArgArgAlaAspAlaValProProGlyArgGlnAlaLeuArgGluArgHis 90
QY 74 AGG-----AAGCCATGGAAGCCCTCAAGGAAACCCACGATTGACCTGGA----- 119
Db ||||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
91 ArgAlaArgProLeuProProCysArgProAlaSerArgArg-ProGlySerSerGI 110
QY 120 -----AAACAGTACGTTCCCTCGGACCAACCAAGGGTATCCGGTTTGATG 169
Db ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
110 YArgHisArgArgLeuLeuHisGlyGlnGlnLeuGln----- 124
QY 170 AACACCCGGAATACCTGCATGACAAAGGTTTGGGCTACGGCCACCTACAACAACGACTTGA 229
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
125 ----ProArgAlaProAlaCysArgGlyArgGly----- 134
QY 230 ACGGGATCAAGACCAACGACATCATGCTGGGTGTCTACATCCCTGACGAAGACGCTCG 289
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
135 -----ProArgGluGluArgProAr 141
QY 290 GCCTGGGCGATGGAACCTGGGTTAGCCCTTGAGCCAAAGCAAGTACGTCCTTTTGGTCATCC 349
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
141 gProGly---AlaThrGlyAsnArgArgArgProValAlaAla----- 154
QY 350 CGGACCAAGACTACGGCAAGCCGATCAACCTCATGAGCTGGGGCGTCAGCGACAACGTGA 409
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
155 -GlyArgArgLeuSerSer-----GlyArgArgSerGlyHi 167
QY 410 TCAAGATGAG 419
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 sHisAspGlu 170

RESULT 11
US-10-641-678-62
; Sequence 62, Application US/10641678
```

```
; Publication No. US20050277172A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Butler, Karlene H.
; APPLICANT: Carlson, Thomas J.
; APPLICANT: Hitz, William D.
; APPLICANT: Stoop, Johan M.
; TITLE OF INVENTION: Plastidic Phosphoglucomutase Genes
; FILE REFERENCE: BB1451 USCIP
; CURRENT APPLICATION NUMBER: US/11/094,586
; CURRENT FILING DATE: 2005-03-30
; PRIOR APPLICATION NUMBER: US 09/906,209
; PRIOR FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: US 60/218,712
; PRIOR FILING DATE: 2000-07-17
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97 & PatentIn Version 3.3
; SEQ ID NO 4
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Zea mays
; US-11-094-586-4

Alignment Scores:
Pred. No.: 5,09 Length: 605
Score: 74.50 Matches: 42
Percent Similarity: 43.26% Conservative: 19
Best Local Similarity: 29.79% Mismatches: 42
Query Match: 8.65% Indels: 38
DB: 7 Gaps: 10

US-10-049-750-13 (1-483) x US-11-094-586-4 (1-605)
QY 69 CTACAGGAAGCCATGGAAGCCCTCAAGAAACCCCAACGATTGACCTGGAAAAACAGCTA 128
Db ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
160 MetSerAlaSerHis---AsnProGlyGlyProAspAsnAspTrpGlyIleLysPheAsn 178
QY 129 CGTCCCCTGCACACCAAGGTATCCGGGTGATGATGAACACCCGGAATACCTGCA 188
Db ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
179 TyrSerSerGlyGlnPro-----Ala-ProGluThrIleTh 190
QY 189 TGACAAAGGTTTGGGCTACGGCCACCTCAACAACGACTTG-----AACGGGATCAAGAC 242
Db ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
190 rAspGlnIle-----TyrGlyAsnThrLeuSerIleSerGluIleLysTh 205
QY 243 CAACGACATCATGTGGGTGTCTACATCCCTGACGAAGAAGACGTGCGCCTCGGATGGA 302
Db ::::: ::::: ::::: ::::: ::::: ::::: :::::
205 rAlaAsp-----IleProAspThrAspLeuSerSerValGlyVal-- 218
QY 303 ACTGGGTTACGCTTGAGCAAGGCAAGTACGTCCTTTTGGTTCATC---CCGGACGAAGA 359
Db ::::: ::::: ::::: ::::: ::::: ::::: :::::
219 -----ValSerTyrGlyAspPheAlaIleGluValIleAspProValSerAs 234
QY 360 CTACGGCAAGCGCATCAACCTCATGAGCTGGGGCGCTCAGCGCAACAACGTCATCAAGATGAG 419
Db ::::: ::::: ::::: ::::: ::::: ::::: :::::
234 pTyr-----LeuGluLeuMet-----GluAsnValPheAspPheG1 246
QY 420 CCAGCTGAAGACTTCAACTTCAACAAGCCGGCTTCCGACTTCTACGAAGGTGCCGTATA 479
Db ::::: ::::: ::::: ::::: ::::: ::::: :::::
246 nLeuIleLysAspLeu---LeuSerArgProAspPheArgPheIlePheAspAlaMetHi 265
QY 480 C 480
Db 265 b 265

RESULT 13
US-11-000-463-455
; Sequence 455, Application US/11000463
; Publication No. US20050266423A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Chen, Rui-hong

; Publication No. US20050277172A1
; GENERAL INFORMATION:
; APPLICANT: Day, Anthony, G.
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Gualfetti, Peter
; APPLICANT: Mitchinson, Colin
; APPLICANT: Neefe, Paulien
; APPLICANT: Sandgren, Mats
; APPLICANT: Shaw, Andrew
; APPLICANT: Stahlberg, Jerry
; TITLE OF INVENTION: Novel Variant Hypocrea jecorina CBH1
; FILE REFERENCE: GC772-3
; CURRENT APPLICATION NUMBER: US/10/641,678
; CURRENT FILING DATE: 2003-08-15
; PRIOR APPLICATION NUMBER: US 60/458,853
; PRIOR FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: US 60/458,696
; PRIOR FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: US 60/456,368
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: US 60/404,063
; PRIOR FILING DATE: 2002-08-16
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PaetSEQ for Windows Version 4.0
; SEQ ID NO 62
; LENGTH: 537
; TYPE: PRT
; ORGANISM: Penicillium janthinellum
; NAME/KEY: VARIANT
; LOCATION: 48, 64
; OTHER INFORMATION: Xaa = Any Amino Acid
; US-10-641-678-62

Alignment Scores:
Pred. No.: 4,4 Length: 537
Score: 75.00 Matches: 31
Percent Similarity: 40.68% Conservative: 17
Best Local Similarity: 26.27% Mismatches: 30
Query Match: 8.71% Indels: 40
DB: 6 Gaps: 6

US-10-049-750-13 (1-483) x US-10-641-678-62 (1-537)
QY 110 TTGACTCGAAACAGCTACGTTCCCTGGACACCAAGGTATCCGGGTGATG 169
Db ::::: ::::: ::::: ::::: :::::
404 MetLeuTyrLeuAspSerThrTyrProThrAsnAlaThrSerThrThrProGlyAla--- 422
QY 170 AACACCCGGATACCTGCATGACAGGTTTGGGCTACGGCCA----- 211
Db ::::: ::::: ::::: ::::: :::::
423 ---LysArgGlyThrCys-----AspIleSerArgArgProAsnThrValGluSerThr 439
QY 212 -----CCTACAAACAGCACTTGA 229
Db ::::: :::::
440 TyrProAsnAlaTyrValIleTyrSerAsnIleLysThrGlyProLeuAsnSerThrPhe 459
QY 230 ACGGATCAAGACCAACGACATCATGCTGGGTGTCTACATCCCTGACGAAGAAGACGTG 289
Db ::::: ::::: ::::: ::::: :::::
460 ThrGlyGly-----ThrThrSer-----SerSerSerThrThrThrThrSer 474
QY 290 GCCTGGGCATGGAACCTGGGTTACGCTTGAGCAAGGCAAGTACGTCCTTTTGGTCATCC 349
Db ::::: ::::: ::::: ::::: :::::
475 -----LysSerThrSerThrSerSerSerSer 483
QY 350 CGGACGAGACTACGCAAGCCGATCAACCTCATGAGCTGGGGCGCTCAGCGACA 403
Db ::::: ::::: ::::: ::::: :::::
484 LysThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 501

RESULT 12
US-11-094-586-4
; Sequence 4, Application US/11094586
; Publication No. US20050273866A1
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 23, 2005, 22:20:36 ; Search time 20.985 Seconds
(without alignments)
4429.126 Million cell updates/sec

Title: US-10-049-750-13
Perfect score: 861
Sequence: 1 gatactaataatgccaaaaa.....acgaaggtgcgtatactaa 483

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgn2 1/USPTO.spool.p/US10049750/runat.23122005.113553.6303/app.query.fasta_1.1742
-DB=PIR -OPT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODES=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10049750 @cgn 1 1.83 @runat.23122005.113553.6303 -NCPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELEXT=7

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	637	74.0	158	JC7522	nucleoside deoxyri
2	150.5	17.5	263	S01360	salivary glue prot
3	142	16.5	1832	T31113	mucin-like glycop
4	136.5	15.9	159	E86686	hypothetical prote
5	134.5	15.6	1161	S57180	probable membrane
6	133.5	15.5	354	T46740	microfilarial shea
7	128	14.9	477	S53362	mucin SAC (clone J
8	125.5	14.6	605	S48940	hypothetical prote
9	124	14.4	510	H84824	En/Spm-like transp
10	121.5	14.1	866	T45462	membrane glycoprot
11	121	14.1	662	A45155	mucin FIM-C.1 - Af
12	121	14.1	797	VG8EX1	glycoprotein X pre
13	116.5	13.5	851	T22696	hypothetical prote
14	116	13.5	232	A60095	larval glue protei

15	114	13.2	867	2	T45463	membrane glycoprot
16	113.5	13.2	770	2	T22808	hypothetical prote
17	111	12.9	279	2	S53363	mucin SAC (clone J
18	110.5	12.8	496	2	E90181	hypothetical prote
19	109.5	12.7	750	2	T42614	probable envelope
20	109	12.7	725	2	A41258	a-agglutinin core
21	108.5	12.6	796	2	T21460	hypothetical prote
22	108	12.5	371	2	S20075	promastigote surfa
23	106.5	12.4	402	2	E86185	hypothetical prote
24	105.5	12.3	327	2	S20074	promastigote surfa
25	105	12.2	328	2	S67570	hypothetical prote
26	105	12.2	815	2	JN0689	glutenin, high-mol
27	104.5	12.1	503	2	S63257	probable membrane
28	104.5	12.1	848	2	S02262	glutenin high mole
29	102.5	11.9	164	2	S53641	mucin SAC - human
30	102.5	11.9	838	1	EEWTHW	glutenin, high mol
31	102	11.8	507	2	T44768	antifreeze glycope
32	102	11.8	660	2	JW0067	chitinase (EC 3.2.
33	102	11.8	815	2	B30843	glutenin high mole
34	102	11.8	1367	1	S48478	glucan 1,4-alpha-g
35	101	11.7	328	2	S01359	salivary glue prot
36	101	11.7	1360	2	T33922	hypothetical prote
37	100.5	11.7	417	2	T24618	hypothetical prote
38	100	11.6	1777	2	T34369	hypothetical prote
39	99.5	11.6	339	2	T25562	hypothetical prote
40	99.5	11.6	379	2	S50125	larval glue protei
41	99.5	11.6	705	2	S18733	glutenin high mole
42	98.5	11.4	217	2	S01358	salivary glue prot
43	98.5	11.4	708	2	T19474	hypothetical prote
44	98.5	11.4	790	2	T34293	hypothetical prote
45	98.5	11.4	798	2	T34248	hypothetical prote

ALIGNMENTS

RESULT 1

JC7522
nucleoside deoxyribosyltransferase (EC 2.4.2.6) II - Lactobacillus helveticus
C:Species: Lactobacillus helveticus
C>Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 09-Jul-2004
C:Accession: JC7522; PC7103
R:Okuyama, K.; Noguchi, T.
Biosci. Biotechnol. Biochem. 64, 2243-2245, 2000
A:Title: Molecular cloning and expression of the nucleoside deoxyribosyltransferase-II 9
A:Reference number: JC7522; MUID:21012342; PMID:11129605
A:Accession: JC7522
A:Molecule type: DNA
A:Residues: 1-158 <OKU>
A:Cross-references: UNIPROT:Q9KWF0; UNIPARC:UPI000015C9E3; DDBJ:AB039914
A:Experimental source: strain ATCC 8018
A:Accession: PC7103
A:Molecule type: protein
A:Residues: 1-11;53-62 <OK>
A:Cross-references: UNIPARC:UPI000017A510; UNIPARC:UPI000017A511
C:Comment: This enzyme catalyzes transfer of glycosyl residues from a donor deoxynucleos

C:Keywords: glycosyltransferase; pentosyltransferase

Alignment Scores:	5.64e-50	Length:	158
Pred. No.:	637.00	Matches:	114
Score:	87.10%	Conservative:	21
Percent Similarity:	73.55%	Mismatches:	20
Best Local Similarity:	73.98%	Indels:	0
Query Match:	2	Gaps:	0
DB:			

US-10-049-750-13 (1-483) x JC7522 (1-158)

QY 16 AAAAAGACGATCTACTTCGGTCCCGCTGGTTCACTGACCGCAAAACAAAGCCTACAAG 75
Db 4 LysLysThrLeuTyrrPheGlyAlaGlyTrpPheAsnGluLysGlnAsnLysalaTyrrLys 23


```
Qy 320 GCCAAGGCAAGTACGTCCTTTGGTCATCCCGGACGAGACTACGGCAAGCGGATCAACC 379
Db 449 LysLysProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 468
Qy 380 TCATGAGCTGGGGCTCAGCGACCAACGATGATCAAGATGAGCGACGCTGAAGGACTTCAACT 439
Db 469 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 480
Qy 440 TCACAAGCCGCGCTTCGACTTACGAAGTGCGGTATACT 481
Db 481 ThrThrThrThrAlaThrThrThrThrLysLysProThrThr 494

RESULT 4
E86686
hypothetical protein yejD [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C:Accession: E86686
R:Boletini, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: E86686
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-159 <STO>
A:Cross-references: UNIPROT:Q9CI73; UNIPARC:UPI00000D43D3; GB:AE005176; PID:g12723374; E
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yejD

Alignment Scores:
Pred. No.: 0.000216 Length: 159
Score: 136.50 Matches: 43
Percent Similarity: 47.83% Conservative: 34
Best Local Similarity: 26.71% Mismatches: 57
Query Match: 15.85% Indels: 27
DB: 2 Gaps: 7

US-10-049-750-13 (1-483) x E86686 (1-159)
Qy 25 ATCTACTTCGGTGGCGGTGTTCTACTGACCGCGCAAAACAAAGCCATACAGGAGCCATG 84
Db 11 ValTyrLeuAlaAlaProPheSerGluSerGlnIleLysLysValGluLeuLeuGlu 30
Qy 85 GAAGCCCTCAAGGAAACCCCAACGATTGACCTGGAAACAGCTAGCTTCCCTCGGACAAAC 144
Db 31 AsnAlaLeuSerLysAsnLysThrVal-----AlaAsnPheSerProMetArgCys 48
Qy 145 CAGTACAAGGGTATCCGGGTGATGAACACCCGGAATACCTGCATGACAAGTT----- 198
Db 49 Gln-----HisProGluSerLeuProGlnGluValGluAla 60
Qy 199 -----TGGCTACGGCCACCTACACACAGCACTGAACGGGATCAAGACCAAC 246
Db 61 PheThrProGluTrpAlaLysAlaThrMetGluAsnAspValAsnGluValAsnLysAla 80
Qy 247 GACATCATGCTGGGTCTACATCCCTGACGAAGAGAGCTC-----GGCCTGGCGCATG 300
Db 81 AspIleIleValAlaIleValAspPheAspHisGlnAspThrAspSerGlyThrAlaTrp 100
Qy 301 GAACCTGGGTACGCTTGAGCAAGCAAGTACGTCCTTTGGTTCATCCCGGACGAAGAC 360
Db 101 GluLeuGlyTyrAlaIleAlaLeuGluLysProThrTyrLeuIleArgPheGluAspThr 120
Qy 361 TAGCGAAGCCGATCAACTCATGAGCTGGGGCTGACGCAAC-----GTG 408
Db 121 Ile-----ProAlaAsnIleMet-----LeuThrGluArgAsnArgAlaPhePhe 135
Qy 409 ATCAAGTACGACGCTGAAGCACTTCACTTCAACAGCCGCGCTTCGACTTCTACGNA 468
Db 136 ThrGlnIleGluGlnValGluGluTyrAspPheLeuGluSerLysLeuIleProTyrSer 155
```

```
Qy 469 GGT 471
Db 156 Gly 156

RESULT 5
S57180
probable membrane protein YJR151c - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein J2223; serine/threonine-rich protein YJR151c
C:Species: Saccharomyces cerevisiae
C>Date: 23-Aug-1995 #sequence_revision 08-Sep-1995 #text_change 09-Jul-2004
C:Accession: S57180
R:Scarcez, T.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S57169
A:Accession: S57180
A:Molecule type: DNA
A:Residues: 1-1161 <SCA>
A:Cross-references: UNIPROT:P47179; UNIPARC:UPI0000128E2A; EMBL:Z49651; NID:g1015902; P1
C:Genetics:
A:Gene: SGD:DNA4; MIPS:YJR151c
A:Cross-references: SGD:S0003912
A:Map position: 10R
C:Keywords: transmembrane protein

Alignment Scores:
Pred. No.: 0.000388 Length: 1161
Score: 134.50 Matches: 47
Percent Similarity: 47.53% Conservative: 30
Best Local Similarity: 29.01% Mismatches: 78
Query Match: 15.62% Indels: 7
DB: 2 Gaps: 3

US-10-049-750-13 (1-483) x S57180 (1-1161)
Qy 5 ACTAAATGCCAAAAAGACGATCTACTTCGGTGGCGGTGTTCTACTGACCGCAAAACA 64
Db 127 ThrThrThrThrLysSerSerThrThrThrProThrThrThrIleThrSerThrThr 146
Qy 65 AAGCCTACA---AGGAAGCCATGGAAGCCCTCAAGGAAACCAACGATTGACCTGGAAA 121
Db 147 SerThrThrSerThrThrProThrThrSerThrThrThrProThrThrSerThr 166
Qy 122 ACAGCTAGTTCCTCCCTGGACAAACAGTACAGGATATCCGGTTCATGACACCCGGAAT 181
Db 167 ThrSerThrThrProThrThrSerThrThrThrProThrThrProThrThrSerThr 186
Qy 182 ACCTGCATGACAAGGTTTGGGCTACGGCCACCTACAACAACGACTTGAACGGGATCAAGA 241
Db 187 ThrThrProThrThrSerThrThrSerThrThrProThrProThrThrSerThrThr 206
Qy 242 CCAACGACATCATGCTGGGTGCTCATCTCCCTGACGAAGAAGAGCTCGGCTGGGCATGG 301
Db 207 ProThrThrSerThrThrSerThrThrProThrThrSerThrThrSerThrThrPro 226
Qy 302 AACTGGGTACGCTTGAGCCCAAGCAAGTACGTCCTTTGGTCA----- 346
Db 227 ThrSerThrThrPro---ThrThrSerThrThrSerThrThrSerGlnThrSerThr 245
Qy 347 TCCCGGACGAAGACTACGGCAAGCGGATCAACCTCATGAGCTGGGGCGCTCAGCGCAACG 406
Db 246 SerThrThrProThrThrSerSerThrThrThrProThrProThrThrSerThrThr 265
Qy 407 TGATCAAGTACGACGCTGAAGGACTTCAACTTCAACAACCGCGCTTCGACTTCTACG 466
Db 266 ThrSerThrThrSerThrAlaProThrThrSerThrThrSerThrThrSerThr 285
Qy 467 AAGGTG 472
Db 286 ThrIle 287

RESULT 6
T46740
microfilarial sheath protein SHP3 [imported] - Litomosoides sigmodontis
```


A:Map position: 8L
C;Keywords: transmembrane protein

Alignment Scores:

Pred. No.: 0.00242 Length: 605
Score: 125.50 Matches: 51
Conservative: 23
Best Local Similarity: 45.96%
Query Match: 31.68% Mismatches: 48
Indels: 39
Gaps: 7
DB:

US-10-049-750-13 (1-483) x S48940 (1-605)

Qy 5 ACTAAATGCCAAAAGAGCATCTACTCGTGGCGGCTGTTCACTGACCGCCAAAACA 64
Db 181 ThrThrThrSerThrLysLeuSerThrSerLeuProThr---SerThrThrSerThr 199
Qy 65 AAGCCTACAAGGAGCCATGGAAGCCCTCAAGGAAACCCCAACGATTGACCTGGAAACA 124
Db 200 SerThrThrThrSerThrSerThrSerThrThrValSerValThrSerThr 219
Qy 125 GCTACGTTCCCTGGACAACGATGACAGGATATCCGGGTTGTGAACACCCGGAATACC 184
Db 220 SerThr---ThrThrSerThrThrSerThrLeuLeSerThr--- 233
Qy 185 TGCATGACAGGTTGGCTAGCGCCACCTACACACGACTTGACCGGATCAAGACCA 244
Db 234 ---SerThrSerSerSerSerThrPro 243
Qy 245 ACGACA---TCATGCTGGGTGCTATACCTCCCTGACGAGAAAGACGCTCGGCTGGGCA 298
Db 244 ThrThrThrSerSerAlaProIleSerThrSerThrSerThrSerThr 261
Qy 299 TGGAACTGGGTACGCTTGAGCCAAAGCAAGTACGCTCTTTTGGTATCCCGACGAAG 358
Db 262 ---SerThrSerThrSerProThrSerSerSerAlaPro 274
Qy 359 ACTAGCGAAGCCGATCAACCTCATGACCTGGGGCTGAGGACACGATGATCAAGATGA 418
Db 275 ThrSerSerSerAsnThr---ProThrSerThrThr 286
Qy 419 GCCAGCTGAAGGACTTCAACTTCAACAGCGCGCTTCGACTTCTACGAAGTGGCGPAT 478
Db 287 -----PheThrThrThrSerProSerThrAlaProSerSerThrThrValThrTyr 303
Qy 479 ACT 481
Db 304 Thr 304

RESULT 9

H84824
En/Spm-like transposon protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C;Accession: H84824
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.;
euss, D.; Niernann, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; PMID:20083487; PMID:10617197
A;Accession: H84824
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-510 <STO>
A;Cross-references: UNIPROT:004210; UNIPARC:UPI00000A9658; GB:AE002093; NID:g2088658; PI
C;Genetics:
A;Gene: At2g40070
A;Map position: 2

Alignment Scores:

Pred. No.: 0.00327 Length: 510
Score: 124.00 Matches: 51

Percent Similarity: 48.68%
Best Local Similarity: 33.55%
Query Match: 14.40%
DB: 2

Conservative: 23
Mismatches: 56
Indels: 22
Gaps: 6

US-10-049-750-13 (1-483) x H84824 (1-510)

Qy 32 TCGGTGCGGCTGGT-----TCACTGACCGCCAAAACAAAGCCCTACAAGGAAGCCA 82
Db 160 AlaThrProThrGlyArgSerSerThrLeuThrAlaAsnSerLysSerSerArg----- 177
Qy 83 TGAAGCCCTCAAGGAAACCCCAACGATTGACCTGGAAAACAGCTACGTTCCCTCGGACA 142
Db 178 -----ProSerThrProThrSerArgAlaThrValSerSerAlaThrArgProSerLeu 195
Qy 143 ACAGTACAAGGATATCCGGTGTATGAACACCCGGGAATACCTGCATGACAAAGTTTGGG 202
Db 196 ThrAsnSerArgSerThrValSerAlaThrThrLysProThrProMetSerArgSerThr 215
Qy 203 CTAGGCCACCTACAACACGACTTGAACGGGATCAAGACCAACGACATCATGCTGGGTG 262
Db 216 SerLeuSerSerSerArgLeuThrProThrAlaSerLysProThrThrSerThrAlaArg 235
Qy 263 TCTACA---TCCCTGACGAAGAAGACGCTCGGCTGGGCATGGAACTGGGTTACGCTTGA 319
Db 236 SerAlaGlySerValThrArgSerThrPro----- 245
Qy 320 GCCAAGCAAGTAGCTCTTTTGGTATCCCGACGAAGACTACGGCAAGCCGATCAACC 379
Db 246 SerThrThrThrLysSerAlaGlyProSerArgSerThrProLeuSerArgSerThr 265
Qy 380 TCATGACTGGGGCTGAGCGACACGCTGATCAAGATGAGCCAGCTGAAGGACTTCAACT 439
Db 266 Ala-----ArgSerSerThrProThrSerArg---ProThrLeuProProSerLys 281
Qy 440 TCAACAAAGCGCGCTTCGACTTCTACGAAGGTGGCG 475
Db 282 ThrIleSerArgSerSerThrProThrArgArgPro 293

RESULT 10

T45462

membrane glycoprotein [imported] - equine herpesvirus 1

C;Species: equine herpesvirus 1

C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004

C;Accession: T45462

R;Kirisawa, R.; Kobayashi, T.; Kawakami, Y.; Iwai, H.

J. Equine Sci. 7, 79-87, 1996

A;Title: Nucleotide sequences of open reading frames 1, 24 and 71 of an attenuated equin

A;Reference number: 222973

A;Accession: T45462

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-866 <KIR>

A;Cross-references: UNIPROT:O39781; UNIPARC:UPI00000ECB1; EMBL:D88733; PIDN:BAA20037.1

A;Experimental source: strain HHI

C;Genetics:

A;Note: ORF1

C;Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homo

F;558-866/Domain: equine herpesvirus 1 glycoprotein homology <EHG>

Alignment Scores:

Pred. No.: 0.00578 Length: 866
Score: 121.50 Matches: 43
Conservative: 25
Best Local Similarity: 41.72%
Query Match: 26.38% Mismatches: 84
Indels: 11
Gaps: 2
DB:

US-10-049-750-13 (1-483) x T45462 (1-866)

Qy 26 TCTACTCGGTGCGGCTGGTTCACTG-----ACCGCAAAACAAAGCCCTACAAGGAAG 79

Db 97 ThrThrSerIleProThrSerThrSerThrGluThrThrThrProThrAlaSer 116

```

QY      80 CCATGGAGCCCTCAGGAAACCCCAACGATTGACCTGGAAACAGCTAGCTTCCCTGG 139
Db      117 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 136
QY      140 ACAACAGTACAAGGTTATCCGGTTGATGACACACCCGGGAATACCTGCATGACAAGTTT 199
Db      137 ThrThrAlaAlaSerThrAlaAlaSerThrSerAlaGluThrThrThrThrThrThrThr 156
QY      200 GGGCTACGGCCACTCAACAACAGCTTGAACGGGATCAAGACCAACGACA----- 250
Db      157 AlaThrSerThrProThrThrThrThrThrProThrProThrSerThrThrThrThrThr 176
QY      251 -----TCATCTGGGTCTCATCCCTGACCTCCCTGACGAGAGAGCTGGCC 292
Db      177 ThrValProThrThrAlaAlaThrThrThrThrThrThrThrThrThrThrThrThrAla 196
QY      293 TGGGCATGGAACCTGGGTTAGCCTTGAGCCCAAGCAAGTACGTCCTTTTGGTCATCCGG 352
Db      197 AlaThrThrThrAlaAlaThrThrThrThrThrThrThrThrThrThrThrThrThrAla 216
QY      353 ACGAAGACTACGGCAAGCCGATCAACCTCATGAGCTGGGCGTCAGCGACAAGCTGATCA 412
Db      217 AlaThrThrThrAlaAlaThrThrThrThrThrAlaAlaThrThrThrThrThrAla 236
QY      413 AGATGACCGACTGAAGGACTCAACTTCAACAGCGCGCTTCGACTTCTACGAAGGTG 472
Db      237 AlaThrThrThrAlaAlaThrThrSerSerAlaThrThrAlaAlaThrThrThrAlaAla 256
QY      473 CCGTATACT 481
Db      257 ThrThrThr 259

RESULT 11
A45155
mucin FIM-C.1 - African clawed frog (fragment)
C:Species: Xenopus laevis (African clawed frog)
C>Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 09-Jul-2004
C:Accession: A45155
R:Hauser, F.; Hoffmann, W.
J. Biol. Chem. 267, 24620-24624, 1992
A:Title: P-domains as shuffled cysteine-rich modules in integumentary mucin C.1 (FIM-C.1)
A:Reference number: A45155; MUID:93077556; PMID:1447205
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-662 <HAU>
A:Cross-references: UNIPROT:Q05049; UNIPARC:UPI000012F8CF; GB:L02115; NID:g214147; PIDN:
F;162-202/Domain: trefoil homology <TRF1>
F;307-347/Domain: trefoil homology <TRF2>
F;354-394/Domain: trefoil homology <TRF3>
F;526-566/Domain: trefoil homology <TRF4>
F;573-613/Domain: trefoil homology <TRF5>
F;621-661/Domain: trefoil homology <TRF6>

Alignment Scores:
Pred. No.: 0.00627 Length: 662
Score: 121.00 Matches: 44
Percent Similarity: 47.52% Conservative: 23
Best Local Similarity: 31.21% Mismatches: 58
Query Match: 14.05% Indels: 16
DB: 2 Gaps: 4

US-10-049-750-13 (1-483) x A45155 (1-662)

QY      47 TCATGACCGCCCAAAACAAAGCCCTACAGGAGCGCTGAGCCCTCAAGAAACCCAA 106
Db      396 SerGlnValAlaAlaThrLysThrThrThrThrThrThrThrThrThrThrThrThrThr 415
QY      107 CGATTGACCTGGAAA---ACAGCTACGTTCCCTCGGCAACCAACGATACAAAGGGTATCCGGG 163
Db      416 ThrThrThrThrLysAlaAlaThrThrThrThrThrThrThrThrThrThrThrThrThr 431

```

```

QY      164 TTGATGAACACCGGAATACCTGCATGACAAAGGTTTGGGCTACGGCCACCTACAAACAG 223
Db      432 -----ThrProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 448
QY      224 ACTTGAACCGGGATCAAGACCAACGACATCATGCTGGGTGCTACATCCCTCGACGAAGAAG 283
Db      449 ThrProThrThrThrThrThrProThrThrThrThrLysAlaAlaThrThrThrThrProThr 468
QY      284 AGTCGGCTGGGCATGGAACCTGGGTTACGCTTGAGCCCAAGCAAGTACGTCCTTTTGG 343
Db      469 ThrThrThr-----ThrThrProThrThrThrThrThrThrLysAlaAlaThrThr 483
QY      344 TCATCCCGGACGAGACTACGGCAAGCCGATCAACCTCATGAGCTGGGCGTCAGCGACA 403
Db      484 ThrThrProThrThrThrThrThrThrThrProThrThrThrThrThrLysAlaAlaThrThr 503
QY      404 ACCTGATCAAGATGAGCCAGCTTCAACCTTCAACAGCGCGCTTCGACTTCT 463
Db      504 ThrProThr-----ThrThrThrThrThrThrThrThrThrThrThrThrLysAlaAlaThrThr 520
QY      464 ACG 466
Db      521 Thr 521

RESULT 12
VGBEX1
glycoprotein X precursor - equine herpesvirus 1 (strain Ab4p)
C:Species: equine herpesvirus 1
A:Note: host Equus caballus (domestic horse)
C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 09-Jul-2004
C:Accession: H36802
R:Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
submitted to GenBank, March 1992
A:Description: The DNA sequence of equine herpesvirus-1.
A:Reference number: A36805
A:Accession: H36802
A:Molecule type: DNA
A:Residues: 1-797 <TEL>
A:Cross-references: UNIPROT:P28968; UNIPARC:UPI0000138750; GB:M86664; NID:g330791; PIDN:
R:Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
Virology 189, 304-316, 1992
A:Title: The DNA sequence of equine herpesvirus-1.
A:Reference number: A41831; MUID:92295586; PMID:1318606
A:Contents: annotation; possible protein-coding frames
A:Note: neither amino acid nor nucleotide sequence is given
C:Genetics:
A:Gene: 71
C:Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homolog
C:Keywords: glycoprotein; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-787/Product: glycoprotein X #status predicted <MAT>
F;23-465/Region: serine/threonine-rich
F;489-797/Domain: equine herpesvirus 1 glycoprotein homology <EHG>
F;766-790/Domain: transmembrane #status predicted <TMN>
F;590/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.: 0.00637 Length: 797
Score: 121.00 Matches: 44
Percent Similarity: 44.65% Conservative: 27
Best Local Similarity: 27.67% Mismatches: 76
Query Match: 14.05% Indels: 12
DB: 1 Gaps: 3

US-10-049-750-13 (1-483) x VGBEX1 (1-797)

QY      26 TCATCTCGTGGCGGTGTTCTACTGACCGCCAAACAAAGCTACAAAGACCCATGG 85
Db      97 ThrThrSerIleProThrSerThrSerThrGluThrThrThrThrThrThr----- 112
QY      86 AAGCCCTCAAGGAAACCCCAACGATGACCTGGAAACAGCT-----ACGTTCCCC 136
Db      113 ---ProThrAlaSerThrThrThrThrThrThrThrThrThrThrAlaAlaProThrThrAlaAla 131

```



```
Db 94 ThrThrThrProCysProThrThr-----ThrThrArgThr 107
Qy 182 ACCTGCATGCAAGGTTTGGGTACGGCCACTCAACAACGACTTGAACGGGATCAAGA 241
Db 108 Thr-----ThrThrProThrThrThrArgThrThrThr 120
Qy 242 CCAACGACATCATGCTGGGTGCTACATCCCTGACGAGAGAGCGCTGGGCATGG 301
Db 121 ProThrThrThrThrArgThrThrThrProThrThrThr 135
Qy 302 AACTGGGTACGCTTGAGCCAAAGCAAGTACGTCCTTTGGTCATCCCGACGAGACT 361
Db 136 -----ArgThrThrThrThrArgThrThrThrProThr 147
Qy 362 ACGCAAGCCGATCAACTCATAGCTGAGCTGGCGCTCAGCGACACGTGATCAAGATGAGCC 421
Db 148 ThrThrThrArgThrThr-----ThrProThrThr 158
Qy 422 AGCTGAAGGACTTCAACTTCAACAGCGCGCTTCGACTTCTACGAAGGTGCGGTACT 481
Db 159 -----ThrThrProThrThrThrGlnThrThrThrArgAlaProProThr 175

RESULT 15
T45463
membrane glycoprotein [imported] - equine herpesvirus 1
C;Species: equine herpesvirus 1
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004
C;Accession: T45463
J;Kiraawa, R.; Kobayashi, T.; Kawakami, Y.; Iwai, H.
J;Equine Sci. 7, 79-87, 1996
A;Title: Nucleotide sequences of open reading frames 1, 24 and 71 of an attenuated equine herpesvirus 1
A;Reference number: Z22973
A;Accession: T45463
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-867 <KIR>
A;Cross-references: UNIPROT:O39782; UNIPARC:UPT00000EF109; EMBL:D88734; PIDN:BAA20038.1
A;Experimental source: isolate 3F clone; strain BK343
C;Genetics:
A;Note: ORF71
C;Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homolog
```

```
Alignment Scores:
Pred. No.: 0.0279 Length: 867
Score: 114.00 Matches: 43
Percent Similarity: 44.03% Conservative: 27
Best Local Similarity: 27.04% Mismatches: 73
Query Match: 13.24% Indels: 16
DB: 2 Gaps: 4
```

US-10-049-750-13 (1-483) x T45463 (1-867)

```
Qy 26 TCTACTTCGGTGGCGGTGGTTCACGCGCCAAACAAAGCCTACAAGAGCCATGG 85
Db 97 ThrThrSerIleProThrSerThrSerThrGluThrThrThr----- 112
Qy 86 AAGCCCTCAAGGAAACCAACGATTCACCTGGAAAAACAGCT-----ACGTTCCCC 136
Db 113 ---ProThrAlaSerThrThrThrThrThrAlaAlaProThrThrAlaAla 131
Qy 137 TGGACAACAGATACAGGGTATCCGGGTGATGATGAACACCCGGAATACCTGCATGCAAGG 196
Db 132 ThrThrThrAlaValThrThrAlaAlaSerThrSerAlaGluThrThrAlaThrAla 151
Qy 197 TTGGGCTACGGCCACCTACACACGACTTGACGGGATCAAGCCAGCATCATGC 256
Db 152 ThrAlaThrSerThrProThrThrThrProThrSerThrThrThrThrAlaThr 171
Qy 257 -----TGGGTGCTTACATCCCTGACGAGAGAGAGCTGGCGCTGGGCATGGAAC 304
Db 172 ThrThrValProThrThrAlaSerThrThrThrThrThrThrThrThrThrThr 191
Qy 305 TGGGTTACGCTTGAGCCAAAGCAAGTACGTCCTTTTGGTCATCCCGGACGAGACTACG 364
```

```
Db 192 AlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThr 211
Qy 365 GCAAGCCGATCAACCTCATGAGCTGGGGCTCAGCGACAAACGTCGATCAAGATGAGCCAGC 424
Db 212 AlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThr-----ThrAla 227
Qy 425 TGAAGGACTTCAACTTCAACAGCCGCGCTTCGACTTCTACGAAGGTGCGGTACTACT 481
Db 228 AlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThr 246
```

Search completed: December 23, 2005, 22:46:45
Job time : 27.985 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 23, 2005, 22:15:36 ; Search time 130.382 Seconds
(without alignments)
5227.239 Million cell updates/sec

Title: US-10-049-750-13
Perfect score: 861
Sequence: 1 gtatacaatgccaaaaa.....acgaaggtgcgtatactaa 483

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool_p/US10049750/runat_23122005.113552.6292/app_query.fasta_1.1742
-DB=UniProt -QFWT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=spct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HAPSIZES=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10049750 @CGN 1 1 614 @runat_23122005.113552.6292 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPOP=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	843	97.9	156	1 NTD_LACLE	Q9r5v5 lactobacill
2	728	84.6	158	2 Q8LY4 LACHE	Q8rly4 lactobacill
3	719	83.5	159	3 Q5FM8 LACAC	Q5fmm8 lactobacill
4	653	75.8	149	1 NTD_LAGJO	Q7alg9 lactobacill
5	634	73.6	158	2 Q9KWF0 LACHE	Q9krf0 lactobacill
6	209	24.3	168	1 NTD_LACFE	Q6yni5 lactobacill
7	171.5	19.9	167	2 Q8LY5 LACHE	Q8rly5 lactobacill
8	154.5	17.9	167	2 Q5FIN0 LACAC	Q5fin0 lactobacill
9	150.5	17.5	263	1 SGS3 DROYA	P13728 drosophila
10	142	16.5	1832	2 Q9S503 CRYPV	Q9s503 cryptospori
11	138	16.0	1286	2 Q9VR49 DROME	Q9vr49 drosophila
12	137.5	16.0	648	2 Q4QJ2 LEIMA	Q4qj2 leishmania
13	136.5	15.9	159	2 Q9CI73 LACLA	Q9ci73 lactococcus
14	136.5	15.9	716	2 Q4P695 USTMA	Q4p695 ustilago ma
15	136.5	15.9	1124	2 Q5CVM4 CRYPV	Q5cvm4 cryptospori
16	136.5	15.9	1126	2 Q5CIS2 CRYHO	Q5cis2 cryptospori

17	136	15.8	1937	2	Q5CV21 CRYPV	Q5cv21 cryptospori
18	135	15.7	668	2	Q4QK4 LEIMA	Q4qk4 leishmania
19	134.5	15.6	1161	1	DAN4 YEAST	P47179 saccharomyc
20	134	15.6	1646	2	Q5CFZ6 CRYHO	Q5cfz6 cryptospori
21	133.5	15.5	354	2	Q25402 LITS1	Q25402 litomosoid
22	133	15.4	881	2	Q8IMS9 DROME	Q8ims9 drosophila
23	132.5	15.4	704	2	Q4QK1 LEIMA	Q4qk1 leishmania
24	132	15.3	519	2	Q7YTR7 CAEEL	Q7ytr7 caenorhabdi
25	130.5	15.2	753	2	Q5SK9 DICDI	Q5sk9 dictyosteli
26	128.5	14.9	916	2	Q7Y210 MONBE	Q7y210 monosiga br
27	128	14.9	333	2	Q6VAV8 ANOGE	Q6vav8 anopheles g
28	128	14.9	1763	2	Q4HX73 GIBZE	Q4hx73 gibberella
29	127	14.8	1349	2	Q8WMO4 HUMAN	Q8wmq4 homo sapien
30	126.5	14.7	1117	2	Q54S26 DICDI	Q54sz6 dictyosteli
31	126	14.6	457	2	Q86AK1 DICDI	Q86aki dictyosteli
32	126	14.6	1795	2	Q76894 DROME	Q76894 drosophila
33	125.5	14.6	605	1	WSC4 YEAST	P38739 saccharomyc
34	125	14.5	334	2	Q6VAV0 ANOGE	Q6vaw0 anopheles g
35	125	14.5	1199	2	Q5CEX1 CRYHO	Q5cex1 cryptospori
36	124.5	14.5	157	2	Q8SL4 LACPL	Q8sl4 lactobacill
37	124	14.4	333	2	Q6VAV7 ANOGE	Q6vav7 anopheles g
38	124	14.4	373	2	Q76810 ANOGE	Q76810 anopheles g
39	124	14.4	510	2	Q04210 ARATH	Q04210 arabidopsis
40	123.5	14.3	374	2	Q8V0L6 9ALPH	Q8v0l6 equid herpe
41	123	14.3	334	2	Q6VAV6 ANOGE	Q6vav6 anopheles g
42	123	14.3	334	2	Q6VAV9 ANOGE	Q6vav9 anopheles g
43	123	14.3	389	2	Q8VOM0 9ALPH	Q8v0m0 equid herpe
44	123	14.3	791	2	Q66VC3 9ALPH	Q66vc3 equid herpe
45	122.5	14.2	669	2	Q5ELV3 DROYA	Q5elv3 drosophila

ALIGNMENTS

RESULT 1
NTD_LACLE
ID NTD_LACLE STANDARD; PRT; 156 AA.
AC Q9R5V5;
DT 25-OCT-2004 (Rel. 45, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 01-FEB-2005 (Rel. 46, Last annotation update)
DE Nucleoside deoxyribosyltransferase (EC 2.4.2.6) (N-
deoxyriboseyltransferase).
GN Name=ntd;
OS Lactobacillus leichmannii.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OX Lactobacillus.
OC NCBI_TaxID=28039;
RN [1]
RP NUCLEOTIDE SEQUENCE. PROTEIN SEQUENCE OF 1-25, FUNCTION, ACTIVE SITE,
RP AND MUTAGENESIS OF GLU-97
RX MEDLINE=95318137; PubMed=7797550; DOI=10.1074/jbc.270.26.15551;
RA Porter D.J.T., Merrill B.M., Short S.A.;
RT "Identification of the active site nucleophile in nucleoside 2-
deoxyriboseyltransferase as glutamic acid 98.";
RL J. Biol. Chem. 270:15551-15556(1995).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RX MEDLINE=96419140; PubMed=8805514;
RA Armetrong S.R., Cook W.J., Short S.A., Balick S.E.;
RT "Crystal structures of nucleoside 2-deoxyriboseyltransferase in native
and ligand-bound forms reveal architecture of the active site.";
RL Structure 4:97-107(1996).
CC -!- FUNCTION: Catalyzes the cleavage of the glycosidic bond of 2'-
deoxyribonucleosides and the transfer of the deoxyriboseyl moiety
to an acceptor purine or pyrimidine base.
CC -!- CATALYTIC ACTIVITY: 2-deoxy-D-riboseyl-base(1) + base(2) = 2-deoxy-
D-riboseyl-base(2) + base(1).
CC -!- BIOPHYSICOCHEMICAL PROPERTIES:
CC pH dependence:
CC Optimum pH is 6.0;
CC -!- PATHWAY: Nucleotide metabolism.
CC -!- SUBUNIT: Homohexamer.
CC -!- SIMILARITY: Belongs to the nucleoside deoxyriboseyltransferase


```
CC family.
CC -1- CAUTION: Was originally (Ref.1) thought to originate from E.coli.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC PDB; 1F8X; X-ray; A/B=1-156.
CC PDB; 1F8Y; X-ray; A/B=1-156.
CC InterPro; IPR007710; N_deoxyrib_trans.
CC Pfam; PF05014; Nuc deoxyrib_tr; 1.
CC 3D-structure; Direct protein sequencing; Nucleotide metabolism;
CC Transferase.
CC INIT MET 0 0
CC ACT SITE 97 97 Nucleophile.
CC MUTAGEN 97 97 E->A: Loss of transferase activity.
CC SQ SEQUENCE 156 AA; 17949 MW; 3A3AEC3FD5B4743B CRC64;

Alignment Scores:
Pred. No.: 1,46e-68 Length: 156
Score: 843.00 Matches: 156
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 97.91% Indels: 0
DB: 1 Gaps: 0

US-10-049-750-13 (1-483) x NTD_LACHE (1-156)
Qy 13 CCAAAAAGAGACTACTTCGGTGGCGGTGTTCTACGACCGCCAAAACAAAGCTAC 72
Db 1 ProlysylthreiletyrPheGlyAlaGlyTrpPheThrAspArgGlnAsnLysAlaTyr 20
Qy 73 AAGGAAGCCATGGAGCCCTCAGGAAACCCACGATTGACCTGGAAACAGCTAGTT 132
Db 21 LysGluAlaMetGluAlaLeuLysGluAsnProthreileAspLeuGluAsnSerTyrVal 40
Qy 133 CCCCTGGACACCAAGTACAAAGGGTATCCGGGTGTATCAACACCCGGGAATACCTGCATGAC 192
Db 41 ProLeuAspAsnGlnTyrLysGlyLeuValAspGluHisProGluTyrLeuHisAsp 60
Qy 193 AAGTTGGGTACGGCCACTACAACACGACTTGAACGGGATCAAGACCAACGACATC 252
Db 61 LysValTrpAlaThrAlaThrTyrAsnAspLeuAsnGlyIleLysThrAsnAspIle 80
Qy 253 ATGCTGGGTCTACATCCCTGACGAGAGACGTCGGCTGGCGCTGGCATGGAAGTGGTAC 312
Db 81 MetLeuGlyValTyrIleProAspGluGluAspValGlyLeuGlyMetGluLeuGlyTyr 100
Qy 313 GCCTTGAGCCAAAGCAAGTACGTCCTTTTGTCTATCCCGACGAAAGACTACGGCAAGCCG 372
Db 101 AlaLeuSerGlnGlyLysTyrValLeuLeuValIleProAspGluAspTyrGlyLysPro 120
Qy 373 ATCAACTCATAGCTGGGGCGTCAGCGACAACTGATCAAGATGACGACCTGAAGGAC 432
Db 121 IleAsnLeuMetSerTrpGlyValSerAspAsnValIleLysMetSerGlnLeuLysAsp 140
Qy 433 TTCACCTTCACAGCCCGCTTCGACTTCTACGAAGTGGCGGTATAC 480
Db 141 PheAsnPheAsnLysProArgPheAspPheTyrGluGlyAlaValTyr 156
```

```
RESULT 2
Q8RLV4_LACHE
ID Q8RLV4_LACHE PRELIMINARY; PRT; 158 AA.
AC Q8RLV4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE N-deoxyribosyltransferase.
GN Nameentd;
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
```

```
OC Lactobacillus.
OX NCBI_TaxID=1587;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CNR232;
RX MEDLINE=21964041; PubMed=11836245; DOI=10.1074/jbc.M111995200;
RA Kaminski P.A.;
RT "Functional cloning, heterologous expression, and purification of two
RT different N-deoxyribosyltransferases from Lactobacillus helveticus.";
RL J. Biol. Chem. 277:14400-14407(2002).
DR EMBL; AY064167; AAL73114.1; -; Genomic_DNA.
DR HSSP; Q8RSV5; 1F8Y.
DR SNR; Q8RLV4; 4-158.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR007710; N_deoxyrib_trans.
DR Pfam; PF05014; Nuc_deoxyrib_tr; 1.
KW Transferase.
SQ SEQUENCE 158 AA; 18148 MW; 00FC9E4B210C47B8 CRC64;

Alignment Scores:
Pred. No.: 5,43e-58 Length: 158
Score: 728.00 Matches: 128
Percent Similarity: 95.48% Conservative: 20
Best Local Similarity: 82.58% Mismatches: 7
Query Match: 84.55% Indels: 0
DB: 2 Gaps: 0

US-10-049-750-13 (1-483) x Q8RLV4_LACHE (1-158)
Qy 16 AAAAAAGAGACTACTTCGGTGGCGGTGTTCTACGACCGCCAAAACAAAGCTACAG 75
Db 4 LysLysThrLeuTyrPheGlyAlaGlyTrpPheAsnGluLysGlnAsnLysAlaTyrLys 23
Qy 76 GAAGCCATGGAAGCCCTCAAGGAAACCCCAACGATTGACCTGGAAACAGCTACGTTCCC 135
Db 24 GluAlaMetAlaLeuLysGluAsnProthreileAspLeuGluAsnSerTyrValPro 43
Qy 136 CTGGACCAACCAAGTATCCGGGTGTATCAACACCCGGGAATACCTGCATGACAG 195
Db 44 LeuGluAsnGlnTyrLysGlyIleArgIleAspGluHisProGluTyrLeuHisAsnIle 63
Qy 196 GTTGGGTACGGCCACTACAACAGACTTGAACGGGATCAAGACCAACGACATCATG 255
Db 64 GluTrpAlaSerAlaThrTyrHisAsnAspLeuValGlyIleLysThrSerAspValMet 83
Qy 256 CTGGGTGTCTACATCCCTGACGAGAAAGACGTCGGCTGGCGCTGGCATGGAAGTGGTACGCC 315
Db 84 LeuGlyValTyrIleProAspGluGluAspValGlyLeuGlyMetGluLeuGlyTyrAla 103
Qy 316 TTGAGCCAAAGCAAGTACGTCCTTTTGTCTATCCCGACGAAAGACTACGGCAAGCCGATC 375
Db 104 LeuSerGlnGlyLysTyrIleLeuLeuValIleProAspGluAspTyrGlyLysProIle 123
Qy 376 AACCTCATAGCTGGGGCGTCAAGCAAGCTGATCAAGATGACCGACCTGAAGGACTTC 435
Db 124 AsnLeuMetSerTrpGlyValCysAspAsnAlaIleLysIleSerGluLeuLysAspPhe 143
Qy 436 AACTTCAACCAAGCCCGCTTCGACTTCTACGAAGTGGCGGTATAC 480
Db 144 AspPheAsnLysProArgTyrAsnPheTyrAspGlyAlaValTyr 158

RESULT 3
Q5FMM8_LACAC
ID Q5FMM8_LACAC PRELIMINARY; PRT; 159 AA.
AC Q5FMM8;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE 2-deoxyribosyltransferase.
GN OrderedLocusNames=LBA0145;
OS Lactobacillus acidophilus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
```

```

OX NCBI_TaxID=1579;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCFM;
RX PubMed=156711160; DOI=10.1073/pnas.0409188102;
RA Altmann E., Russell W.M., Azcarate-Peril M.A., Barrangou R.,
RA Luck B.L., McAlliffe O., Southern N., Dobson A., Duong T., Callanan M.,
RA Lick S., Hamrick A., Cano R., Klaenhammer T.R.;
RT "Complete genome sequence of the probiotic lactic acid bacterium
RT Lactobacillus acidophilus NCFM.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:3906-3912(2005).
DR EMBL; CP000033; AAV42046.1; -; Genomic DNA.
GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR007710; N_deoxyrib_tr; 1.
DR Pfam; PF05014; N_deoxyrib_tr; 1.
KW Complete proteome; Transferase.
SQ SEQUENCE 159 AA; 18252 MW; 5CCTFFDBA57938FE CRC64;

Alignment Scores:
Pred. No.: 3 65e-57 Length: 159
Score: 719.00 Matches: 128
Percent Similarity: 94.81% Conservative: 18
Best Local Similarity: 83.12% Mismatches: 8
Query Match: 83.51% Indels: 0
DB: 2 Gaps: 0

US-10-049-750-13 (1-483) x Q5FM08_LACAC (1-159)
QY 19 AAGACGATCTACTTCGGTCCGGCTGGTTCACTGACCGCCAAACAAAGCCTACAAGGAA 78
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 6 LysThrLeuTyrPheGlyAlaGlyTrpPheAsnGluLeuGlnAsnLysAlaTyrLysAla 25
QY 79 GCATGGAAAGCCCTCAGGAAACCAACGAGTACCTGGAAACAGACTACGTTCCCTG 138
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 26 AlaMetGluAlaLeuLysGlnAsnProThrValAspLeuGluAsnSerTyrValProLeu 45
QY 139 GACACACGATACAGGGTATCGGGTGGTATGACACCGCGGAATACCTGCATGACAGGTT 198
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 46 GluAsnGlnTyrLysAspIleArgValAspGluHisProGluTyrLeuHisAspIleGlu 65
QY 199 TGGGCTACGGCCACCTTACCAACACGACTTGAACGGGATCAAGCAACCAACGACATCATGCTG 258
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 66 TrpAlaSerAlaThrTyrHisAsnAspLeuIleGlyLeuSerSerAspIleMetLeu 85
QY 259 GTGTCTACATCCCTGACGAAGACGTCGGCTGGGATCGATGGAAGTTCAGCCTTG 318
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 86 GlyValTyrLeuProGluGluAspValGlyLeuGlyMetGluLeuGlyTyrAlaLeu 105
QY 319 AGCCAGGCAAGTACGTCCTTTTGGTTCATCCCGGACGAGACTACGGCAAGCCGATCAAC 378
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 106 SerGlnGlyLysTyrIleLeuLeuValIleProAspGluAspTyrGlyLysProIleAsn 125
QY 379 CTCATGAGCTGGGGCTCAGCACAACGTCATGATCAAGATCAGCGAGCTGAAGGACTTCAAC 438
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 126 LeuMetSerTrpGlyValCysAspAsnAlaIleLysIleSerGluLeuLysAspPheAsp 145
QY 439 TTCAACAGCCGGCTTCGACTTCTACGAAGTGGCGTATAC 480
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 146 PheAsnLysProArgPheAsnPheTyrAspGlyAlaValTyr 159

RESULT 4
NTD_LACJO
ID NTD_LACJO STANDARD; PRT; 149 AA.
AC Q74LQ9;
DT 25-OCT-2004 (Rel. 45, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Nucleoside deoxyribosyltransferase (EC 2.4.2.6) (N-
DE deoxyribosyltransferase).
GN Name:ntd; OrderedLocusNames=J0124;
OS Lactobacillus johnsonii.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.

```

```

Db      142 AsnPhetYrAspGlyAlaValTyr 149
      :::|||||:::|||||
RESULT 5
Q9KWF0 LACHE
ID Q9KWF0_LACHE PRELIMINARY; PRT; 158 AA.
AC Q9KWF0
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nucleoside deoxyribosyltransferase-II.
GN Name=ndtB;
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1587;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21012342; PubMed=11129605; DOI=10.1271/bbb.64.2243;
RA Okuyama K., Noguchi T.;
RT "Molecular cloning and expression of the nucleoside
RT deoxyribosyltransferase-II gene from Lactobacillus helveticus.";
RL Biosci. Biotechnol. Biochem. 64:2243-2245(2000).
DR EMBL; AB039914; BAA92683.2; -; Genomic_DNA.
DR PIR; JC7522; JC7522.
DR HSP; Q9RSV5; IF8Y.
DR SMR; Q9KWF0; 4-158.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR007710; N deoxyrib trans.
DR Pfam; PF05014; Nuc_deoxyrib_tr; 1.
KW Transferase.
SQ SEQUENCE 158 AA; 18318 MW; D661273AD89309CD CRC64;

Alignment Scores:
Pred. No.: 2,37e-49 Length: 158
Score: 634.00 Matches: 113
Percent Similarity: 87.10% Conservative: 22
Best Local Similarity: 72.90% Mismatches: 20
Query Match: 73.64% Indels: 0
DB: 2 Gaps: 0

US-10-049-750-13 (1-483) x Q9KWF0_LACHE (1-158)
Qy 16 AAAAAGCATCTACTTCGGTGGCGGTGTTCACTGACCGCCAAAACAAAGCCTACAAG 75
|:::|||||:::|||||
Db 4 LysLysThrLeuTyrPheGlyAlaGlyTrpPheAsnGluLysGlnAsnLysAlaTyrLys 23
|:::|||||:::|||||
Qy 76 GAAGCCATGGAGCCCTCAGGAAACCCACGATTGACCTGGAAACAGCTACGTTCCC 135
|:::|||||:::|||||
Db 24 GluAlaMetAlaAlaLeuLysGluAsnProThrValAspLeuGluAsnSerTyrValPro 43
|:::|||||:::|||||
Qy 136 CTGGACAACAGTACAAAGGGTATCCGGTGTGATCAACACCGGAATACCTGCATGACAAG 195
|:::|||||:::|||||
Db 44 LeuGluAsnGlnTyrLysGlyIleArgIleAspGluHisProGlnTyrLeuHisAsnIle 63
|:::|||||:::|||||
Qy 196 GTTTGGGCTACGGCCACTACAACAGCTTGTAACGGGATCAAGACAACGACATCATG 255
|:::|||||:::|||||
Db 64 GluTrpAlaSerAlaThrTyrHisAsnAspLeuValGlyIleLysThrSerAspValLeu 83
|:::|||||:::|||||
Qy 256 CTGGGTGTACATCCCTGACGAGAAGACAGTCCGGCTGGCATGGAACTGGGTAGCC 315
|:::|||||:::|||||
Db 84 LeuGlyValTyrLeuProGlnGluHisValGlyLeuGlyMetGluLeuGlyTyrPro 103
|:::|||||:::|||||
Qy 316 TTGAGCAAGCAAGTAGTACGTCCTTTTGTCTATCCGGACGAAGACTACGGCAAGCCGATC 375
|:::|||||:::|||||
Db 104 LeuSerGlnGlyLysLeuPhePheTrpPheSerHisMetLysAspTyrGlyLysProIle 123
|:::|||||:::|||||
Qy 376 AACCTCATGAGCTGGGGCGTCAGCGACAAACGTGATCAAGATGAGCCAGCTGAAGACTTC 435
|:::|||||:::|||||
Db 124 IleLeuMetSerTrpGlyValCysAspAsnAlaSerGlnIleSerGluLeuLysAspPhe 143
|:::|||||:::|||||
Qy 436 AACTTCACAGCCGGCTTCGATCTTACGAAGTCCGTTATAC 480
|:::|||||:::|||||

Db      144 AspPheAsnLysProArgTyrAsnPhetYrAspGlyAlaValTyr 158
      :::|||||:::|||||
RESULT 6
NTD_LACFE
ID NTD_LACFE STANDARD; PRT; 168 AA.
AC Q6YNI5;
DT 25-OCT-2004 (Rel. 45, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Nucleoside deoxyribosyltransferase (EC 2.4.2.6) (N-
DE deoxyribosyltransferase).
GN Name=ntd;
OS Lactobacillus fermentum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1613;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=CIP 102980T;
RA Kaminski P.A.;
RT "Characterization of the Lactobacillus fermentum N-
RT deoxyribosyltransferase.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Catalyzes the cleavage of the glycosidic bond of 2'-
CC deoxyribonucleosides and the transfer of the deoxyribosyl moiety
CC to an acceptor purine or pyrimidine base (By similarity).
CC -!- CATALYTIC ACTIVITY: 2-deoxy-D-ribose(1) + base(2) = 2-deoxy-
CC D-ribose(1) + base(1).
CC -!- PATHWAY: Nucleotide metabolism.
CC -!- SIMILARITY: Belongs to the nucleoside deoxyribosyltransferase
CC family.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; AY064168; AAL73115.1; -; Genomic_DNA.
CC InterPro; IPR007710; N deoxyrib trans.
CC Pfam; PF05014; Nuc_deoxyrib_tr; 1.
CC Nucleotide metabolism; Transferase.
KW Nucleotide metabolism; Transferase.
FT ACT_SITE 103 103 Nucleophile (By similarity).
SQ SEQUENCE 168 AA; 18896 MW; 362162F43586C317 CRC64;

Alignment Scores:
Pred. No.: 2,73e-10 Length: 168
Score: 209.00 Matches: 55
Percent Similarity: 51.66% Conservative: 23
Best Local Similarity: 36.42% Mismatches: 59
Query Match: 24.27% Indels: 14
DB: 1 Gaps: 7

US-10-049-750-13 (1-483) x NTD_LACFE (1-168)
Qy 25 ATCTACTTCGGTCCGCTGGTTCACCTGACCCGCAAAACAAAGCCTACAAGGAGCCATG 84
|:::|||||:::|||||
Db 12 IleTyrLeuAlaThrSerPhePheAsnGluGluArgAlaArgIleProGlnAlaLeu 31
|:::|||||:::|||||
Qy 85 GAAGCCCTCAAGAAACCCACGATTGACCTGGAAACAGCTACGTTCCCTGGACAAC 144
|:::|||||:::|||||
Db 32 AlaGlnLeuGluAlaAsnProThrValGlyValHisGln-----ProPheAspPhe 49
|:::|||||:::|||||
Qy 145 CAGTACAAGGTATCCGGGTGTGATGAACACCCGAATACCTGCATGACAAGGTT---TGG 201
|:::|||||:::|||||
Db 50 GlnTyrLysAspAlaArgValAspSerAspProAlaGlyValPheGlySerLeuGluTrp 69
|:::|||||:::|||||
Qy 202 GCTACGGCCACCTACAACACGACTTGAACGGGATCAAGACCAACGACATCATGCTGGGT 261
|:::|||||:::|||||
Db 70 GlnIleAlaThrTyrAsnAsnAlaValGlyThrSerAspValCysValAla 89
|:::|||||:::|||||
Qy 262 GTCTACATCCCTGACGAGAGAGCGTCCGCTGGGATGGAAGCTGGGTACGCTTGAGC 321
|:::|||||:::|||||
```

```

Db 90 LeuTyrAspMetAspGlnIleAspGluGlyIleCysMetGluIleGlyMetPheValAla 109
Qy 322 CAAGCAAG---TAGTCTCTTTTGGTTCATCCGGACGGAAGACTCGGCAAG---CCGATC 375
Db 110 LeuHisLysProIleValLeuLeuProPheThrLysLysAspLysSerAlaTyrGluAla 129
Qy 376 AACCTCATG-----AGCTGGGGCGTCAGCGACACAGCTGATCAAGATG 417
Db 130 AnLeuMetLeuAlaArgGlyValThrTrp---LeuGluProAsn-----AspPhe 146
Qy 418 AGCCAGCTGAAGGACTTCAACTTCAACCAAGCCG 450
Db 147 SerProLeuLysAspPheAsnHisPro 157

RESULT 7
Q8RLY5_LACHE
ID Q8RLY5_LACHE PRELIMINARY; PRT; 167 AA.
AC Q8RLY5_LACHE PRELIMINARY; PRT; 167 AA.
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Purine trans deoxyribosylase (Nucleoside deoxyribosyltransferase-I)
DE (EC 2.4.2.6).
GN Name=ptd; Synonyms=ndtA;
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI TaxID=1587;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CNR232;
RX MEDLINE=21964041; PubMed=11836245; DOI=10.1074/jbc.M111995200;
RA Kaminski P.A.;
RT "Functional cloning, heterologous expression, and purification of two
different N-deoxyribosyltransferases from Lactobacillus helveticus.";
RL J. Biol. Chem. 277:14400-14407(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Okuyama K., Noguchi T.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AV064166; AAL73113.1; -; Genomic DNA.
DR EMBL; AB076265; BAC00952.1; -; Genomic DNA.
DR FDB; 182D; X-ray; A/B/Cs-.
DR PDB; 182G; X-ray; A/B/Cs-1-167.
DR PDB; 182I; X-ray; A/B/Cs-1-167.
DR PDB; 182L; X-ray; A/B/Cs-1-167.
DR PDB; 183F; X-ray; A/B/Cs-1-167.
DR GO; GO:0050144; P.nucleoside deoxyribosyltransferase activity; IEA.
DR GO; GO:0016757; P.transferase activity, transferring glycosyl. . .; IEA.
DR InterPro; IPR007710; N_deoxyrib_trans.
DR Pfam; PF05014; Nuc_deoxyrib_tr; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 167 AA; 18713 MW; 5540581511CB4787 CRC64;

```

Alignment Scores:

Pred. No.:	7.64e-07	Length:	167
Score:	171.50	Matches:	49
Percent Similarity:	53.21%	Conservative:	34
Best Local Similarity:	31.41%	Mismatches:	58
Query Match:	19.92%	Indels:	15
DB:	2	Gaps:	7

```

US-10-049-750-13 (1-483) x Q8RLY5_LACHE (1-167)
Qy 10 ATGCCAAAAGACGACTTCTTCGTCGGCTGCTTCTACCTGACGCCCAAAACAAAGCC 69
Db 5 ValProThrGlyLysIleTyrLeuGlySerProPheTyrSerAspAlaGlnArgGluArg 24
Qy 70 TACAAGGAAGCCATGGAAGCCCTCAAGGAAAACCCCAAGATTGACCTGGAAAACAGCTAC 129
Db 25 AlaAlaLysAlaLysGluLeuLeuAlaLysAsnProSerIle-----AlaHisValPhe 42
Qy 130 GTTCCCTCGGACCAACGACTACAAAGGGTATCCGGGTGTGATGAACACCCGGAAATAC----- 183

```

```

Db 43 PheProPheAspGlyPheThrAsp---ProAspGluLysAsnProGluIleGlyGly 61
Qy 184 CTGCATGACAGGTTTGGGTACGCGCACCTACCAACACGACTTGAACGGGATCAAGACC 243
Db 62 IleArgSerMetValTrpArgAspAlaThrTyrGlnAsnAspLeuThrGlyIleSerAsn 81
Qy 244 AACGACATCATGCTGGGTGCTACATCCCTCAGCAAGAAGACGTCGGCCTGGCATCGAA 303
Db 82 AlaThrCysGlyValPheLeuTyrAspMetAspGlnLeuAspAspGlySerAlaPheGlu 101
Qy 304 CTGGTTTACCGCTTGAGCAAGCAAGTACGTCTCTTTTGGTC-----ATCCCG 351
Db 102 IleGlyPheMetArgAlaMetHisLysProValIleLeuValProPheThrGluHisPro 121
Qy 352 GACGAAGACTACGCAAGCCGATCAACTCATG---AGCTGGGGCGTCAGC----- 399
Db 122 GluLysGlu-----LysLysMetAsnLeuMetIleAlaGlnGlyValThrThrIleIle 139
Qy 400 GACAACTGATCAAGATGAGCCAGCTGAGGACTTCAAGACTTCAACTTCAACAG 447
Db 140 AspGlyAsnThrGluPheGluLeuAlaAspTyrAsnPheAsnGlu 155

RESULT 8
Q5FINO_LACAC
ID Q5FINO_LACAC PRELIMINARY; PRT; 167 AA.
AC Q5FINO;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Putative deoxyribosyltransferase.
DE OrderedLocusNames=LBA1631;
GN Lactobacillus acidophilus.
OS Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI TaxID=1579;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCFM;
RX PubMed=15671160; DOI=10.1073/pnas.0409188102;
RA Altermann E., Russell W.M., Azcarate-Peril M.A., Barrangou R.,
RA Buck B.L., McAuliffe O., Souther N., Dobson A., Duong T., Callanan M.,
RA Lick S., Hamrick A., Cano R., Klesenhammer T.R.;
RT "Complete genome sequence of the probiotic lactic acid bacterium
Lactobacillus acidophilus NCFM.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:3906-3912 (2005).
DR EMBL; CP000033; AAV43444.1; -; Genomic DNA.
DR GO; GO:0016740; P.transferase activity; IEA.
DR InterPro; IPR007710; N_deoxyrib_trans.
DR Pfam; PF05014; Nuc_deoxyrib_tr; 1.
KW Complete proteome; Transferase.
SQ SEQUENCE 167 AA; 18933 MW; AB0408F06C3E4D98 CRC64;

```

Alignment Scores:

Pred. No.:	2.79e-05	Length:	167
Score:	154.50	Matches:	49
Percent Similarity:	49.01%	Conservative:	25
Best Local Similarity:	32.45%	Mismatches:	60
Query Match:	17.94%	Indels:	17
DB:	2	Gaps:	7

```

US-10-049-750-13 (1-483) x Q5FINO_LACAC (1-167)
Qy 25 ATCTACTTCGGTCGGCTGTTCTACTGACCGCCAAAACAAAGCCTACAAGGAAGCCATG 84
Db 10 IleTyrLeuGlyThrProPheTyrAsnAspGlnArgAlaArgValLysLysAlaArg 29
Qy 85 GAAGCCCTCAAGGAAAACCCCAAGATTGACCTGGAAAACAGCTAGTTCCTCCCTGGACAC 144
Db 30 AlaLeuLeuGluGlnAsnProThrVal-----ValArgValHisPheProPheAspGln 47
Qy 145 CAGTACAAAGGGTATCCGGGTGAT-----GAACACCCGGAATAC-----CTGCAT 189

```



```

Db      31 AsnAlaLeuSerLysAsnLysThrVal-----AlaAsnPheSerProMetArgCys 48
QY      145 CAGTCAAGGTTATCCGGGTTGATGAACACCCGGAATACCTGCATGACAAGTT-----198
Db      49 Gln-----HisProGluSerLeuProGlnGluValGluAla 60
QY      199 -----TGGGCTACGGCCACCTACACACACGACTTGACGGGATCAAGCCAC 246
Db      61 PheThrProGluTrpAlaLysAlaThrMetGluAsnAspValAsnGluValAsnLysAla 80
QY      247 GACATCATGCTGGGGTGTCTACATCCTCGACGAAGAAGACGTC-----GGCTGGGCATG 300
Db      81 AspIleIleValAlaIleValAspPheAspHisGlnAspThrAspSerGlyThrAlaTrp 100
QY      301 GAATGGGTTTACGCTTGAGCGAAGCAAGTACGTCCTTTGGTTCATCCCGGACGAAGAC 360
Db      101 GluLeuGlyTyrAlaIleAlaLeuGluLysProThrTyrLeuIleArgPheGluAspThr 120
QY      361 TAGCGAAGCGATCAACTCATGAGCTGGGGCGTCAGCGACAAC-----GTG 408
Db      121 Ile-----ProAlaAsnIleMet-----LeuThrGluArgAsnArgAlaPhePhe 135
QY      409 ATCAAGATGAGCGACGCTGAGGACTTCAACTTCAACAGCGCGCTTCGACTTCTACGAA 468
Db      136 ThrGlnIleGluGlnValGluGluIuTrpAspPheLeuGluSerLysLeuIleProTyrSer 155
QY      469 GGT 471
Db      156 Gly 156

RESULT 14
Q4P695_USTMA
ID Q4P695 USTMA PRELIMINARY; PRT; 716 AA.
AC Q4P695
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=UM04368.1;
OS Ustilago maydis 521.
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_TaxID=237631;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=521;
RA Birren B., Nuebea C., Abebe A., Abouelleil A., Adekoya E.,
RA Ait-zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,
RA Arachchi H., Armbruster J., Bachantsang P., Baldwin J., Barry A.,
RA Bayul T., Blitshteyn B., Bloom T., Blye J., Boguslavskiy L.,
RA Borowsky M., Boukhgalter B., Brunsche A., Butler J., Calixte N.,
RA Calvo S., Camarata J., Campo K., Chang J., Cheehatsang Y., Citroen M.,
RA Collymore A., Considine T., Cook A., Cooke P., Corum B., Cuomo C.,
RA David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,
RA Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engels R.,
RA Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA Fitzgerald M., Foley K., Gage D., Galagan J., Gearin G., Gnerre S.,
RA Gnikre A., Goyette A., Graham J., Grandbois E., Gvaltsen K., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,
RA Jaffe D., Jones C., Kamal M., Kamat A., Kamysselis M., Karlsson E.,
RA Kelle C., Kieu A., Kiener P., Kodira C., Kulbokas E., Labutti K.,
RA Lama D., Lenders T., Leger J., Levine S., Lewis D., Lewis T.,
RA Lindblad-toh K., Liu X., Lokyitsang T., Lokyitsang Y., Lucien O.,
RA Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,
RA Manning J., Maraballa R., Maru K., Matthews C., Mauceli E.,
RA McCarthy M., McDonough S., Mcghee T., Meidrim J., Mensu L.,
RA Mesirov J., Mihalev A., Mihova T., Mikelsen T., Menga V., Moru K.,
RA Mozes J., Mulrain L., Munson G., Naylor J., News C., Nguyen C.,
RA Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari M., Norbu C.,
RA Norbu N., O'donnell P., Okoawo O., O'leary S., Omosho B.,
RA O'Neill K., Osman S., Parker S., Perrin D., Phunkhang P., Piquani B.,

```

```

RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,
RA Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA Rutman M., Schupbach R., Seaman C., Settippalli S., Sharpe T.,
RA Sheridan J., Shierpa N., Shi J., Smirnov S., Smith C., Sougnéz C.,
RA Spencer B., Stalker J., Stange-thomann N., Stavropoulos S.,
RA Stetson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA Tenzing P., Tesfaye S., Theodore J., Thoulutsang Y., Topham K.,
RA Towey S., Tsamla T., Tsomo N., Vallee D., Vassiliev H.,
RA Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadvav S.,
RA Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
RA Zimmer A., Zody M., Lander E.;
RT "The genome sequence of Ustilago maydis.";
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AACP01000154; EAK85643.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 716 AA; 76918 MW; COA3C76508C98321 CRC64;

Alignment Scores:
Pred. No.: 0.00153 Length: 716
Score: 136.50 Matches: 54
Percent Similarity: 50.00% Conservative: 26
Best Local Similarity: 33.75% Mismatches: 57
Query Match: 15.85% Indels: 23
DB: 2 Gaps: 7

US-10-049-750-13 (1-483) x Q4P695_USTMA (1-716)
QY 2 TATACTAAATGCCAAAAA-----AGACGATCTACTTCGGTCCGGCTGTTCACTGACC 55
Db 166 TyrArgArgCysThrGlnPheValArgAlaThr-----177
QY 56 GCCAAAAACAAAGCCTACAAAGGAGCCATGGAAGCCCTCAAGGAAAC-----CAACA 109
Db 178 -----ThrThrSerArgArgLysThrThrSerThrLysArgLysThrSerThrLys 195
QY 110 TTGACCTGGAAAAACAGCTACGCTTCCCTTGGACACACAGTACAAAGGATCCGGGTTGATG 169
Db 196 ArgThrSerThrThrThrSerThrArgArgThrThrSerThrSerLysArgThrThr 215
QY 170 AACACCCCGGAATACCTGATGACAAAGTTGGGCTACGGCCACCTACAAACAGCACTGA 229
Db 216 SerThrLeuAsnAlaValArgThrAlaSerSerThrArgArgSerThrThrThrSer 235
QY 230 ACGGATCAAGCAACCAACGACATCATGCTGGGTGTCTACATCCTCACCAGGAAGACGTCG 289
Db 236 ThrThrProLysProSerThrThrSer-----ThrThrSerArgSerSerThrThrSer 253
QY 230 GCCTGGGCATGGAACCTGGGTTACGCTTGGACCAAGCAAGTACGCTCTTTTGGTCATCC 349
Db 254 LysProSerThrSerSerThrThr-----SerArgSerSerThrAlaSerThrThrSer 271
QY 350 CGGACGAAGACTACGGCA---AGCGATCAACCTCATGAGCTGGGGCGCTACCGCAACG 406
Db 272 ArgSerSerThrThrSerThrSerThrSerThrSerSerSerThrAsnArgAlaLeuThrThr 291
QY 407 TGATCAAGATGAGCCAGCTGAGGACTTCAACTTCAACAAAGCCGCTTCGACTTCTAGC 466
Db 292 -----SerSerSerArgSerSerThrSerThrThrThrSerSerSerThrThrThr 307

RESULT 15
Q5CW4_CRYPV PRELIMINARY; PRT; 1124 AA.
ID Q5CW4_CRYPV
AC Q5CW4
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Secreted protein with cysteine rich repeats and a mucin like threonine
DE rich repeat, signal peptide, transcripts identified by EST.
GN ORFNames=cgd8_3520;

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2005, 21:47:05 ; Search time 188 Seconds
(without alignments)
366.928 Million cell updates/sec

Title: US-10-049-750-14
Perfect score: 848
Sequence: 1 MPKXTIYFGAGWFTDRQNKA.....QLKDFNFKNKPRDFYEGAVY 157

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A. Geneseq 21.*
1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*
9: geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848	100.0	157	4	AAB67589 Amino aci
2	842	99.3	157	8	ADR97103 N-deoxyri
3	728	85.8	158	6	ABP98175 Amino aci
4	719	84.8	159	6	ABP98180 Amino aci
5	637	75.1	158	5	AAG66170 L. helvet
6	629	74.2	133	6	ABP98179 Amino aci
7	405	47.8	84	6	ABP98178 Amino aci
8	209	24.6	168	6	ABP98177 Amino aci
9	166.5	19.6	167	6	ABP98176 Amino aci
10	136.5	16.1	159	5	ABB53802 Lactococc
11	96	11.3	1149	5	ABB77620 AnEPV P4a
12	94	11.1	750	1	AAP82940 Atrial na
13	91.5	10.8	156	7	ADC96860 E. faeciu
14	91	10.7	719	9	ADW47705 F. magna
15	86.5	10.2	462	5	ABB54050 Lactococc
16	85.5	10.1	438	4	AAB95642 Human pro
17	85.5	10.1	456	8	ADH41469 Novel hum
18	85.5	10.1	464	3	AAB23624 Human sec
19	85.5	10.1	487	8	ADH41499 Novel hum
20	85.5	10.1	487	8	ADH41497 Novel hum
21	85.5	10.1	487	8	ADH41493 Novel hum
22	85.5	10.1	487	8	ADH41495 Novel hum
23	85.5	10.1	487	8	ADH41463 Novel hum
24	85.5	10.1	487	8	ADS88243 Human pro

25	84.5	10.0	243	8	ADH41485	Novel hum
26	84.5	10.0	247	8	ADH41467	Novel hum
27	84.5	10.0	396	4	AAB90548	Human sec
28	84.5	10.0	396	5	ABG65427	Human alb
29	84.5	10.0	396	8	ADL78694	Albumin f
30	84.5	10.0	452	8	ADH41489	Novel hum
31	84.5	10.0	456	8	ADH41471	Novel hum
32	84.5	10.0	456	8	ADH41465	Novel hum
33	84.5	10.0	456	8	ADH41477	Novel hum
34	84.5	10.0	456	8	ADH41479	Novel hum
35	84.5	10.0	487	4	AAU08684	Human FCT
36	84.5	10.0	487	4	AAU08683	Human FCT
37	84.5	10.0	487	7	ADB32033	Human FCT
38	84.5	10.0	487	7	ADB32035	Human FCT
39	84.5	10.0	487	8	ADH41491	Novel hum
40	84.5	10.0	493	8	ADH41473	Novel hum
41	84	9.9	706	6	ABO53100	Human put
42	84	9.9	1042	9	ADX07993	Cyclin-de
43	84	9.9	1046	7	ADE58906	Human Pro
44	84	9.9	1046	7	ADD45534	Human Pro
45	84	9.9	1046	7	ADE61258	Human Pro

ALIGNMENTS

RESULT 1
AAB67589
ID AAB67589 standard; protein; 157 AA.

AC AAB67589;

DT 29-MAY-2001 (first entry)

DE Amino acid sequence of a nucleoside 2-deoxyribosyltransferase.

KW Deoxyribonucleoside; deoxyribose 1-phosphate; thymidine phosphorylase;
KW purine nucleoside phosphorylase; phosphopentose mutase;
KW phosphopentose aldolase; fructose 1,6-diphosphate aldolase;
KW deoxyribokinase; nucleoside 2-deoxyribosyltransferase.

OS Lactobacillus leichmannii.

XX WO200114566-A2.

PD 01-MAR-2001.

PF 18-AUG-2000; 2000WO-EP008088.

PR 20-AUG-1999; 99EP-00116425.

XX (HOFF) ROCHE DIAGNOSTICS GMBH.

PA (INSP) INST PASTEUR.

XX (PHAR-) PHARMA-WALDHOF GMBH & CO KG.

PI Tischer W, Ihlenfeldt H, Barzu O, Sakamoto H, Pistotnik E;

PI Marliere P, Pochet S;

DR WPI; 2001-235026/24.

DR N-PSDB; AAF55445.

PT In vitro enzymatic synthesis of deoxyribonucleosides comprises reacting

PT deoxyribose 1-phosphate and a nucleobase to form a deoxyribonucleoside

PT and an inorganic phosphate.

PS Disclosure; Page 62-63; 73pp; English.

XX The present sequence represents a nucleoside 2-deoxyribosyltransferase

CC enzyme. This enzyme is involved in the biosynthesis of

CC deoxyribonucleosides, and is used in the method of the invention. The

CC specification describes a method for the in vitro enzymatic synthesis of

CC deoxyribonucleosides. The method comprises reacting deoxyribose 1-

CC phosphate and a nucleobase to form a deoxyribonucleoside and an inorganic

CC phosphate. Enzymes which may be used in the method of the invention
 CC include thymidine phosphorylase, purine nucleoside phosphorylase,
 CC phosphopentose mutase, phosphopentose aldolase, fructose 1,6-diphosphate
 CC aldolase, deoxyribokinase, and nucleoside 2-deoxyribosyltransferase
 XX
 SQ Sequence 157 AA;

Query Match 100.0%; Score 848; DB 4; Length 157;
 Best Local Similarity 100.0%; Pred. NO. 5.8e-84;
 Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPKKTIFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNQYKGRVDEHPEYLH 60
 DB 1 MPKKTIFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNQYKGRVDEHPEYLH 60
 QY 61 DKVWATATYNNDLNGIKTNDIMLGVIYPDEEDVGLGMEIGYALSOGKYVLLVIPDEDYDK 120
 DB 61 DKVWATATYNNDLNGIKTNDIMLGVIYPDEEDVGLGMEIGYALSOGKYVLLVIPDEDYDK 120
 QY 121 PINLMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157
 DB 121 PINLMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157

RESULT 2
 ADR97103
 ID ADR97103 standard; protein; 157 AA.
 XX
 AC ADR97103;
 XX
 DT 16-DEC-2004 (first entry)
 XX
 DE N-deoxyribosyltransferase (DTP) with G9S mutant, seq id 1.
 XX
 KW Cytosatic; antimicrobial; N-deoxyribosyltransferase; DDTP;
 KW N-deoxyribosyltransferase; DTP; antitumour; infection; ddc; ddt; mutein.
 XX
 OS Lactobacillus leichmannii.
 XX
 OS Synthetic.
 XX
 PH Key Location/Qualifiers
 FT Misc-difference 9 /note= "Wild-type Gly replaced with Ser"
 FT
 XX
 XX FR2852968-A1.
 XX
 XX 01-OCT-2004.
 XX
 XX 28-MAR-2003; 2003FR-00003910.
 XX
 XX 28-MAR-2003; 2003FR-00003910.
 XX
 XX (INSP) INST PASTEUR.
 XX (CNRS) CNRS CENT NAT RECH SCI.
 XX
 XX Kaminski PA;
 XX
 XX WPI; 2004-693069/68.
 XX N-PSDB; ADR97104.
 XX
 XX Method for evolution of proteins to modify characteristics, useful
 PT particularly for evolving an N-deoxyribosyltransferase to produce an N-
 PT dideoxyribosyltransferase for preparation of antitumor nucleosides and
 PT nucleotides.
 XX
 XX Claim 15; SEQ ID NO 2; 51pp; French.
 XX
 XX The invention relates to a method for evolution of a protein (X) so as to
 CC modify its characteristics. Further disclosed are mutated proteins (X*)
 CC produced by the new method and with altered activity, relative to (X).
 CC Also disclosed is a nucleic acid (I) comprising a reference for N-
 CC dideoxyribosyltransferase (DDTP), obtained from an N-
 CC deoxyribosyltransferase (DTP) by using the new method. The method is

CC specifically used to evolve an N-deoxyribosyltransferase (DTP) to an N-
 CC dideoxyribosyltransferase (DDTP), i.e. an enzyme able to transfer
 CC dideoxyribose from one dideoxyribonucleotide to another. Such an enzyme
 CC is used for preparation of nucleotides and nucleosides that have
 CC antitumour activity or can be used to treat infections, particularly ddc
 CC and ddt. The current sequence represents the N-deoxyribosyltransferase
 CC (DTP) amino acid sequence containing a G9S mutation. This protein
 XX displays the activity of N-dideoxyribosyltransferase.
 SQ Sequence 157 AA;

Query Match 99.3%; Score 842; DB 8; Length 157;
 Best Local Similarity 99.4%; Pred. NO. 2.6e-83;
 Matches 156; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPKKTIFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNQYKGRVDEHPEYLH 60
 DB 1 MPKKTIFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNQYKGRVDEHPEYLH 60
 QY 61 DKVWATATYNNDLNGIKTNDIMLGVIYPDEEDVGLGMEIGYALSOGKYVLLVIPDEDYDK 120
 DB 61 DKVWATATYNNDLNGIKTNDIMLGVIYPDEEDVGLGMEIGYALSOGKYVLLVIPDEDYDK 120
 QY 121 PINLMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157
 DB 121 PINLMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157

RESULT 3
 ABP98175
 ID ABP98175 standard; protein; 158 AA.
 XX
 AC ABP98175;
 XX
 DT 11-AUG-2003 (first entry)
 XX
 DE Amino acid sequence of N-deoxyribosyltransferase protein.
 XX
 KW N-deoxyribosyltransferase; ntd gene; enzyme; antitumour chemotherapy;
 KW deoxyribonucleotide synthesis; infection; insecticide; herbicide.
 XX
 OS Lactobacillus helveticus.
 XX
 XX WO2003025163-A2.
 XX
 XX 27-MAR-2003.
 XX
 XX 12-SEP-2002; 2002WO-FR003120.
 XX
 XX 14-SEP-2001; 2001FR-00011911.
 XX
 XX (INSP) INST PASTEUR.
 XX (NARE-) INST NAT RECH AGRONOMIQUE.
 XX
 XX Kaminski P, Tailliez P, Marliere P, Quenee P, Cotaya R;
 XX WPI; 2003-313351/30.
 XX N-PSDB; ACC43580, ACC43586.
 XX
 XX New N-deoxyribosyltransferase from Lactobacillus, useful for preparing
 PT deoxyribonucleotides, e.g. for use as antiviral agents, also related
 PT nucleic acid and antibodies.
 XX
 XX Claim 1; Page 61; 70pp; French.
 XX
 XX The present sequence represents a N-deoxyribosyltransferase protein from
 CC Lactobacillus. The enzyme is encoded by the ntd gene. The N-
 CC deoxyribosyltransferase protein is useful to raise specific antibodies,
 CC and for in vitro or in vivo enzymatic synthesis of deoxyribonucleotides,
 CC particularly those containing non-natural bases. These
 CC deoxyribonucleotides are useful for treating bacterial, viral (e.g. HIV),
 CC parasitic or fungal infections, for antitumour chemotherapy, and as
 CC insecticides or herbicides

```
XX SQ Sequence 158 AA;
Query Match 85.8%; Score 728; DB 6; Length 158;
Best Local Similarity 82.6%; Pred. No. 7.3e-71;
Matches 128; Conservative 20; Mismatches 7; Indels 0; Gaps 0;

Qy 3 KKTIFYGAGWFTDRQNKAYKEAMEALKENPTIDLENSVPLDNOYKGI RVDHEPPEYLHDK 62
Db 4 KKTIFYGAGWFTDRQNKAYKEAMEALKENPTIDLENSVPLDNOYKGI RVDHEPPEYLHNI 63
Qy 63 VWATATYNDNLNGIKTNDIMLGVIIPDEEDVGLGMEGLYALSQGYKYLIVIPDEYDGKPI 122
Db 64 EWASATYHNDLVGIKTSVDMLGVIIPDEEDVGLGMEGLYALSQGYKYLIVIPDEYDGKPI 123
Qy 123 NLMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157
Db 124 NLMSGVCDNAIKISELKDFNFKPRYNYFDGAVY 158

RESULT 4
ID ABP98180 standard; protein; 159 AA.
XX AC ABP98180;
XX DT 11-AUG-2003 (first entry)
XX DE Amino acid sequence of N-deoxyribosyltransferase protein.
XX KW N-deoxyribosyltransferase; ntd gene; enzyme; antitumour chemotherapy;
XX OS deoxyribonucleotide synthesis; infection; insecticide; herbicide.
XX PN Lactobacillus acidophilus.
XX WO2003025163-A2.
XX PD 27-MAR-2003.
XX PF 12-SEP-2002; 2002WO-FR003120.
XX PR 14-SEP-2001; 2001FR-00011911.
XX PA (INSP ) INST PASTEUR.
XX PA (NARE-) INST NAT RECH AGRONOMIQUE.
XX PI Kaminski P, Tailliez P, Marliere P, Quenee P, Cotaya R;
XX WPI; 2003-313351/30.
XX DR N-PSDB; ACC43585.
XX PT New N-deoxyribosyltransferase from Lactobacillus, useful for preparing
XX PT deoxyribonucleotides, e.g. for use as antiviral agents, also related
XX PT nucleic acid and antibodies.
XX PS Claim 1; Page 68; 70pp; French.
XX CC The present sequence represents a N-deoxyribosyltransferase protein from
XX CC Lactobacillus. The enzyme is encoded by the ntd gene. The N-
XX CC deoxyribosyltransferase protein is useful to raise specific antibodies,
XX CC and for in vitro or in vivo enzymatic synthesis of deoxyribonucleotides,
XX CC particularly those containing non-natural bases. These
XX CC deoxyribonucleotides are useful for treating bacterial, viral (e.g. HIV),
XX CC parasitic or fungal infections, for antitumour chemotherapy, and as
XX CC insecticides or herbicides
XX SQ Sequence 159 AA;
Query Match 84.8%; Score 719; DB 6; Length 159;
Best Local Similarity 83.1%; Pred. No. 7e-70;
Matches 128; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

Qy 4 KTIYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSVPLDNOYKGI RVDHEPPEYLHDKV 63
```

```
Db 6 KTIYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSVPLDNOYKGI RVDHEPPEYLHDIE 65
Qy 64 WATATYNDNLNGIKTNDIMLGVIIPDEEDVGLGMEGLYALSQGYKYLIVIPDEYDGKPIN 123
Db 66 WASATYHNDLVGIKTSVDMLGVIIPDEEDVGLGMEGLYALSQGYKYLIVIPDEYDGKPIN 125
Qy 124 LMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157
Db 126 LMSGVCDNAIKISELKDFNFKPRYNYFDGAVY 159

RESULT 5
ID AAG66170 standard; protein; 158 AA.
XX AC AAG66170;
XX DT 17-JUN-2002 (first entry)
XX DE L. helveticus nucleoside deoxyribosyltransferase-II.
XX KW Nucleoside deoxyribosyltransferase-II; enzyme; deoxynucleoside; drug;
XX OS recombinant; ntdB.
XX PN Lactobacillus helveticus.
XX JP2002051781-A.
XX PD 19-FEB-2002.
XX PF 08-AUG-2000; 2000JP-00239443.
XX PR 08-AUG-2000; 2000JP-00239443.
XX PA (YAMS ) YAMASA SHOYU KK.
XX WPI; 2002-299403/34.
XX DR N-PSDB; ABL40579.
XX PT Enzymatic preparation of deoxynucleoside.
XX PS Claim 1; Page 10; 14pp; Japanese.
XX CC The present invention relates to a recombinant nucleoside
XX CC deoxyribosyltransferase-II. The nucleoside deoxyribosyltransferase-II and
XX CC the encoding gene are useful in an enzymatic preparation of a
XX CC deoxynucleoside. The deoxynucleoside is useful as the raw material for
XX CC various drugs. The present sequence represents the L. helveticus
XX CC recombinant nucleoside deoxyribosyltransferase-II (ntdB)
XX SQ Sequence 158 AA;
Query Match 75.1%; Score 637; DB 5; Length 158;
Best Local Similarity 73.5%; Pred. No. 6.2e-61;
Matches 114; Conservative 21; Mismatches 20; Indels 0; Gaps 0;

Qy 3 KKTIFYGAGWFTDRQNKAYKEAMEALKENPTIDLENSVPLDNOYKGI RVDHEPPEYLHDK 62
Db 4 KKTIFYGAGWFTDRQNKAYKEAMEALKENPTIDLENSVPLDNOYKGI RVDHEPPEYLHNI 63
Qy 63 VWATATYNDNLNGIKTNDIMLGVIIPDEEDVGLGMEGLYALSQGYKYLIVIPDEYDGKPI 122
Db 64 EWASATYHNDLVGIKTSVDMLGVIIPDEEDVGLGMEGLYALSQGYKYLIVIPDEYDGKPI 123
Qy 123 NLMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157
Db 124 ILMSGVCDNASQISELKDFNFKPRYNYFDGAVY 158

RESULT 6
ID ABP98179 standard; protein; 133 AA.
```

```

XX AC ABP98179;
XX DT .11-AUG-2003 (first entry)
XX DE Amino acid sequence of N-deoxyribosyltransferase protein.
XX KW N-deoxyribosyltransferase; ntd gene; enzyme; antitumour chemotherapy;
XX KW deoxyribonucleotide synthesis; infection; insecticide; herbicide.
XX OS Lactobacillus amylovorus.
XX XX WO2003025163-A2.
XX PD 27-MAR-2003.
XX PF 12-SEP-2002; 2002WO-FR003120.
XX PR 14-SEP-2001; 2001FR-00011911.
XX XX (INSP ) INST PASTEUR.
XX PA (NARE-) INST NAT RECH AGRONOMIQUE.
XX PI Kaminski P, Tailliez P, Marliere P, Quenee P, Cotaya R;
XX WPI; 2003-313351/30.
XX DR N-PSDB; ACC43584.
XX PT New N-deoxyribosyltransferase from Lactobacillus, useful for preparing
XX PT deoxyribonucleotides, e.g. for use as antiviral agents, also related
XX PT nucleic acid and antibodies.
XX PS Claim 1; Page 66-67; 70pp; French.
XX CC The present sequence represents a N-deoxyribosyltransferase protein from
XX CC Lactobacillus. The enzyme is encoded by the ntd gene. The N-
XX CC deoxyribosyltransferase protein is useful to raise specific antibodies,
XX CC and for in vitro or in vivo enzymatic synthesis of deoxyribonucleotides,
XX CC particularly those containing non-natural bases. These
XX CC deoxyribonucleotides are useful for treating bacterial, viral (e.g. HIV),
XX CC parasitic or fungal infections, for antitumour chemotherapy, and as
XX CC insecticides or herbicides
XX SQ Sequence 133 AA;
Query Match 74.2%; Score 629; DB 6; Length 133;
Best Local Similarity 84.2%; Pred. No. 3.6e-60;
Matches 112; Conservative 17; Mismatches 4; Indels 0; Gaps 0;
Qy 25 MEALKENPTIDLENSYVPLDNQYKGIKRVDEHPHYLHDKVWATATYNNDLNGIKTNDIMLG 84
Db 1 MEALKKNPTVDLENSYVPLDNQYKGIKRVDEHPHYLHDKVWATATYNNDLNGIKSSDVMLG 60
Qy 85 VYLPDEEDVGLGMLGVALSGKVKVLLVLPDEDYGKPINLMSGVSDNVIMKSQLKDFNF 144
Db 61 VYLPDEEDVGLGMLGVALSGKVKVLLVLPDEDYGKPINLMSGVSDNVIMKSELKDFDF 120
Qy 145 NKPRFDYEGAVY 157
Db 121 NRPRNFYDGAVY 133
RESULT 7
ABP98178
ID ABP98178 standard; protein; 84 AA.
XX AC ABP98178;
XX DT .11-AUG-2003 (first entry)
XX DE Amino acid sequence of N-deoxyribosyltransferase protein.
XX KW N-deoxyribosyltransferase; ntd gene; enzyme; antitumour chemotherapy;

```

```

KW XX deoxyribonucleotide synthesis; infection; insecticide; herbicide.
OS XX Lactobacillus crispatus.
PN XX WO2003025163-A2.
XX PD 27-MAR-2003.
XX PF 12-SEP-2002; 2002WO-FR003120.
XX PR 14-SEP-2001; 2001FR-00011911.
XX XX (INSP ) INST PASTEUR.
XX PA (NARE-) INST NAT RECH AGRONOMIQUE.
XX PI Kaminski P, Tailliez P, Marliere P, Quenee P, Cotaya R;
XX WPI; 2003-313351/30.
XX DR N-PSDB; ACC43583.
XX PT New N-deoxyribosyltransferase from Lactobacillus, useful for preparing
XX PT deoxyribonucleotides, e.g. for use as antiviral agents, also related
XX PT nucleic acid and antibodies.
XX PS Claim 1; Page 65; 70pp; French.
XX CC The present sequence represents a N-deoxyribosyltransferase protein from
XX CC Lactobacillus. The enzyme is encoded by the ntd gene. The N-
XX CC deoxyribosyltransferase protein is useful to raise specific antibodies,
XX CC and for in vitro or in vivo enzymatic synthesis of deoxyribonucleotides,
XX CC particularly those containing non-natural bases. These
XX CC deoxyribonucleotides are useful for treating bacterial, viral (e.g. HIV),
XX CC parasitic or fungal infections, for antitumour chemotherapy, and as
XX CC insecticides or herbicides
XX SQ Sequence 84 AA;
Query Match 47.8%; Score 405; DB 6; Length 84;
Best Local Similarity 86.9%; Pred. No. 5.1e-36;
Matches 73; Conservative 8; Mismatches 3; Indels 0; Gaps 0;
Qy 45 NOYKGIKRVDEHPHYLHDKVWATATYNNDLNGIKTNDIMLGVIYPDEDYGLGMLGVALS 104
Db 1 NOYKGIKRVDEHPHYLHDKVWATATYNNDLNGIKSSDMLGVLPDEDYGLGMLGVALS 60
Qy 105 OGKVVLLVLPDEDYGKPINLMSG 128
Db 61 KGKIVLLVLPDEDYGKPINLMSG 84
RESULT 8
ABP98177
ID ABP98177 standard; protein; 168 AA.
XX AC ABP98177;
XX DT 11-AUG-2003 (first entry)
XX DE Amino acid sequence of N-deoxyribosyltransferase protein.
XX KW N-deoxyribosyltransferase; ntd gene; enzyme; antitumour chemotherapy;
XX KW deoxyribonucleotide synthesis; infection; insecticide; herbicide.
XX OS Lactobacillus fermentum.
XX PN WO2003025163-A2.
XX PD 27-MAR-2003.
XX PF 12-SEP-2002; 2002WO-FR003120.
XX PR 14-SEP-2001; 2001FR-00011911.
XX

```


CC present sequence represents the amino acid sequence of the F. magna
CC antibody light chain binding protein, protein L.
SQ Sequence 719 AA;

Query Match 10.7%; Score 91; DB 9; Length 719;
Best Local Similarity 28.0%; Pred. No. 2.1;
Matches 37; Conservative 17; Mismatches 50; Indels 28; Gaps 6;

Qy 11 GWFTDRQNKAYKEAMEALKENP--TIDLENSYVPLDNQYKQIRVDEHPE----- 57
Db 349 GTFPAEATAEAYRYADLLAKENGKYTADLEDGGYTINIRFAGKVDKPEKEQVTIKENI 408

Qy 58 YLHDKVWATATYNNDL-----NGIKTNDIML---GVYIPDEEDVGLGMELGYALSQKQYV 109
Db 409 YFEDGTQVATATFKGTFPAEATAEAYRYADLLSKHGKGYTADLEDGGYTINIRFA---GKEE 465

Qy 110 LLVIPDEDYGKP 121
Db 466 -----PEETPEKP 473

RESULT 15
ABB54050
ID ABB54050 standard; protein; 462 AA.
XX AC ABB54050;
XX AC
XX 29-AUG-2003 (revised)
DT 16-MAY-2002 (first entry)
XX
XX Lactococcus lactis protein yheD.
DE
XX
XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX
XX Lactococcus lactis; IL1403.
XX
XX FR2807446-A1.
XX 12-OCT-2001.
XX
XX 11-APR-2000; 2000FR-00004630.
XX
XX 11-APR-2000; 2000FR-00004630.
XX
XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX
XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX
XX WPI; 2002-043418/06.
XX
XX New nucleotide sequence useful in the identification or Lactococcus
XX lactis and related species.
XX
XX Claim 6; SEQ ID NO 752; 2504pp; French.
XX

The present invention is related to a Lactococcus lactis nucleotide
sequence (ABA90521) and related proteins (ABB53300-ABB55621). The nucleic
acid sequence is useful in the detection and/or amplification of nucleic
acid sequence, particularly to identify Lactococcus lactis or related
species. The proteins of the invention are useful for the biosynthesis or
biodegradation of a composition of interest. The invention helps research
in lactic bacteria, particularly useful in the production of yogurt and
cheese. Note: The sequence data for this patent is based on equivalent
patent WO200177334 (published 18-OCT-2001) which is available in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 29-AUG-2003 to
CC standardise OS field)
XX
XX Sequence 462 AA;

Query Match 10.2%; Score 86.5; DB 5; Length 462;
Best Local Similarity 24.1%; Pred. No. 3.4;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2005, 21:56:48 ; Search time 47 Seconds

(without alignments)
276.172 Million cell updates/sec

Title: US-10-049-750-14

Perfect score: 848

Sequence: 1 MPKTYIFGAGWFTDRQNK.....QLKDFNFKPRDFYEGNAVY 157

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5 COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/6 COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/H COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/pCTUS COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/RE COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91.5	10.8	156	2	US-09-107-532A-6487
2	84.5	10.0	396	2	US-09-800-729-86
3	83.5	9.8	382	1	US-08-477-451-28
4	83.5	9.8	1781	1	US-08-477-451-11
5	82.5	9.7	766	4	PCT-US94-00198-6
6	82.5	9.7	920	2	US-09-248-796A-17529
7	80.5	9.5	2237	1	US-08-354-973-1
8	78.5	9.3	170	2	US-09-134-000A-4783
9	76.5	9.0	338	2	US-09-489-039A-11745
10	76	9.0	279	2	US-09-134-001C-4707
11	76	9.0	434	1	US-08-795-475-3
12	76	9.0	434	1	US-08-325-278B-3
13	75.5	8.9	181	1	US-08-307-499-58
14	75.5	8.9	181	2	US-09-299-268-58
15	75.5	8.9	537	2	US-09-248-796A-19446
16	75	8.8	160	2	US-09-107-532A-4152
17	74	8.7	171	2	US-10-085-572-7
18	74	8.7	171	2	US-10-087-195-7
19	74	8.7	190	2	US-09-618-919A-18
20	74	8.7	750	2	US-09-248-796A-18005
21	74	8.7	993	2	US-08-836-687B-30
22	74	8.7	1646	2	US-09-535-008-67
23	74	8.7	1647	2	US-09-535-008-2
24	74	8.7	1647	2	US-09-824-574-4
25	74	8.7	1647	2	US-09-538-092-1172
26	74	8.7	1649	2	US-09-535-008-75
27	74	8.7	1650	2	US-09-535-008-71

Sequence 9752, Ap
Sequence 69, Appl
Sequence 65, Appl
Sequence 77, Appl
Sequence 73, Appl
Sequence 1948, Ap
Sequence 4434, Ap
Sequence 6783, Ap
Sequence 47, Appl
Sequence 45, Appl
Sequence 28646, A
Sequence 7334, Ap
Sequence 27095, A
Sequence 4391, Ap
Sequence 192, App
Sequence 192, App
Sequence 11595, A
Sequence 99, Appl

28 74 8.7 1659 2 US-09-949-016-9752
29 74 8.7 1678 2 US-09-535-008-69
30 74 8.7 1679 2 US-09-535-008-85
31 74 8.7 1681 2 US-09-535-008-77
32 74 8.7 1682 2 US-09-535-008-73
33 73.5 8.7 252 2 US-09-710-279-1948
34 73.5 8.7 308 2 US-09-134-001C-4434
35 73.5 8.7 499 2 US-09-328-352-6783
36 73.5 8.7 844 2 US-09-422-936-47
37 73.5 8.7 960 2 US-09-422-936-45
38 73 8.6 225 2 US-09-252-991A-28646
39 73 8.6 264 2 US-09-489-039A-7334
40 73 8.6 992 2 US-09-252-991A-27095
41 72.5 8.5 164 2 US-09-107-532A-4391
42 72.5 8.5 331 2 US-09-370-838-192
43 72.5 8.5 331 2 US-09-854-133-192
44 72.5 8.5 631 2 US-09-949-016-11595
45 72.5 8.5 659 2 US-10-152-886-99

ALIGNMENTS

RESULT 1
US-09-107-532A-6487
; Sequence 6487, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-8277
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6487:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...156
; SEQUENCE DESCRIPTION: SEQ ID NO: 6487:

US-09-107-532A-6487

Query Match 10.8%; Score 91.5; DB 2; Length 156;
Best Local Similarity 23.0%; Pred. No. 0.021;
Matches 35; Conservative 28; Mismatches 74; Indels 15; Gaps 5;
QY 6 IYFAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNOYKGIKIRVDEHPYLDKVA 65
DB 14 IYLAGPFSEQIDRVSIRIKALEENKTIV--TSFYSRPHQ-----ESNYELFSAGWA 64
QY 66 TATYNNDLNGIKTNDIMLGVIYIPDEE--DVGLGMLGYALSQGYVLLVIPDEDYDKPIN 123
DB 65 QEVYKDMBELTNAEFVVAILDPEHQIDPSTAYELGVATLKKPMIIV---QEETVPTN 121
QY 124 LMSWGVSDNVKMSQ--LKDFNFKPRDFYEG 154
DB 122 LMITQSLHTLYLKSQDAVREYDFETLPVETVYG 153
RESULT 2
US-09-800-729-86
; Sequence 86, Application US/09800729
; Patent No. 6605592
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 32 Human secreted proteins
; FILE REFERENCE: PZ044P1
; CURRENT APPLICATION NUMBER: US/09/800,729
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: PCT/US00/26013
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 60/155,709
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 86
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-729-86

Query Match 10.0%; Score 84.5; DB 2; Length 396;
Best Local Similarity 23.1%; Pred. No. 0.52;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;
QY 9 GAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNOYK-----GIRVDEHPYLDK 63
DB 251 GDRWILTAHTIYPKDSVSLRKQSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNE- 309
QY 64 WATATYNNDL-----NGIKTNDIMLGVIYIPDEEDV-----GLGMLGYALSQK 107
DB 310 --SHNFGDIALLELQHSILPGPNVLPCLPDNETLYRSLGLGYVSGFGMGWLTTELK 367
QY 108 YVLL-VIPDE 116
DB 368 YSLPLVAPRE 377

RESULT 3
US-08-477-451-28
; Sequence 28, Application US/08477451
; Patent No. 5928665
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori Cagi Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-477-451-28
Query Match 9.8%; Score 83.5; DB 1; Length 382;
Best Local Similarity 26.0%; Pred. No. 0.64;
Matches 38; Conservative 22; Mismatches 71; Indels 15; Gaps 5;
QY 17 ONKAYKEAMEALKENPTID--LENSYVPLDNOYKGIKIRVD-EHPEYLDKVAWATATYNNDL 73
DB 111 KSKAYGDLQALADNPYKLLNPYAVYLNQETFTKEKERSLYYPPQVKTSSEIKTT 170
QY 74 NGIKTND---IMLGVIYIPDEEDVGLGMLGYALSQGYVLLVIPDEDYDKPI-NLMSWGV 129
DB 171 ATTKRAQALLQMGVFSLDEEQNKASRLALSQKA-----IBEYSNNISNLSRKE 222
QY 130 SDNVKMSQLKDFNFKPRDFYEGA 155
DB 223 LDNIYYQLERNKFDKAKDIAQA 248

RESULT 4
US-08-477-451-11
; Sequence 11, Application US/08477451
; Patent No. 5928665
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori Cagi Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 11:

Qy	6	IYFAG----	WFTDRONKAYKAMEALKNENFTIDILNENSYVPLDNOYKGRVDEHPVLHDK	62
		: : : : :	: : : : :	
Db	15	IFPASCAMTWTFTIDNKK-----	YDNRHYTKTINSKIEHL--S	49
		: : : : :	: : : : :	
Qy	63	VWATATYNNDLNGIKTNDIMLGVYIPDEYGL-----	GMELGYAL-----	103
		: : : : :	: : : : :	
Db	50	ISTVTNNIISGKK-----	JAVFTGDKINVTNNKRLSIKEKAVDRGCGYGLNPNPH	104
		: : : : :	: : : : :	
Qy	104	SQCKYVLIVPDEYDKCPINLMSW-----	GVS---DNVI--KMSQLKDFNFNK	146
		: : : : :	: : : : :	
Db	105	SNRKLTIWPEKDL-KSLNIQSLGELDINQVNLKHVSLETDRITQLKESLNQVNI	ES	163
		: : : : :	: : : : :	
Qy	147	PRPDFY	152	
		: : : : :	: : : : :	
Db	164	SKANFY	169	

RESULT 11
 US-08-795-475-3
 ; Sequence 3, Application US/08795475
 ; Patent No. 5965390
 ; GENERAL INFORMATION:
 ; APPLICANT: Bjvrck, Lars
 ; APPLICANT: SJvdring, Ulf
 ; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED AND BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/795,475
 ; FILING DATE: 11-FEB-1997
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mcmasters, David D.
 ; REGISTRATION NUMBER: 33,963
 ; REFERENCE/DOCKET NUMBER: 100084.402D1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 434 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; HYPOTHEetical: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Escherichia coli LE392/pHDLG, DSM 7055
 ; US-08-795-475-3

RESULT 12
 US-08-325-278B-3
 ; Sequence 3, Application US/08325278B
 ; Patent No. 6822075
 ; GENERAL INFORMATION:
 ; APPLICANT: Bjorck, Lars
 ; Sjöberg, Ulf
 ; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed IP Law Group
 ; STREET: 701 Fifth Avenue Suite 6300
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/325.278B
 ; FILING DATE: 26-Oct-1994
 ; CLASSIFICATION: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Potter, Jane E. R.
 ; REGISTRATION NUMBER: 33,332
 ; REFERENCE/DOCKET NUMBER: 100084.402
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 434 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Escherichia coli LE392/pHDLG, DSM 7055
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
 ;

RESULT 13
US-08-307-499-58
; Sequence 58, Application US/08307499
; Patent No. 5651972
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Vi uela, Eladio
; APPLICANT: Gibbs, E.P.J.
; TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a
; TITLE OF INVENTION: Live Vaccine Vector
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik


```
; LENGTH: 537
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-19446

Query Match      8.9% Score 75.5; DB 2; Length 537;
Best Local Similarity 26.5%; Pred. No. 8.7;
Matches 41; Conservative 18; Mismatches 59; Indels 37; Gaps 10;

Qy  2 PKKTIYFGAGW-FTDRQNKAYKEAMEALKENPTIDLENSYVPLDNQYKGIRVDEHPEYLH 60
Db  314 FKPTDEF--LWEFIDKCGK-YK---SPKTNP-LDFYKTYVRFD-----ASEHFSLIH 358

Qy  61 D-----KWTATATYNDLNGIKTNDIMLGV-----YIPDEEDVGLGMELGYALSQ 105
Db  359 DPRNEYNKLTYTVERLNNIFGGKPIEYINLEIDEIKQVAIRKMKONEPVFFGSDVG-XFSD 417

Qy  106 GKYVLLVIPDEDYGKPINLMSWGSVDNVIKMSQLK 140
Db  418 SKSGILDTTAYDYS-----TAFDFSLDITKLQRLK 447
```

Search completed: December 23, 2005, 22:05:47
Job time : 50 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2005, 22:04:24 ; Search time 165 Seconds
(without alignments)
397.571 Million cell updates/sec

Title: US-10-049-750-14
Perfect score: 848
Sequence: 1 MPKXITFGAGWFTDRQNK.....QLKDFNFKPRDFYEGAVY 157

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA Main:
1: /cgm2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgm2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgm2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgm2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgm2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgm2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85.5	10.1	464	3	US-09-808-602-95
2	85.5	10.1	487	3	US-09-808-602-93
3	85.5	10.1	487	3	US-09-800-198-80
4	84.5	10.0	396	3	US-09-800-729-86
5	84.5	10.0	396	3	US-09-833-245-2176
6	84.5	10.0	452	3	US-09-808-602-21
7	84.5	10.0	487	3	US-09-808-602-17
8	84.5	10.0	487	3	US-09-808-602-19
9	84.5	10.0	487	3	US-09-800-198-17
10	84.5	10.0	487	3	US-09-800-198-19
11	84	9.9	1537	6	US-11-097-143-27759
12	84	9.9	1634	5	US-10-732-923-8723
13	84	9.9	1638	5	US-10-732-923-8724
14	84	9.9	1638	6	US-11-097-143-12630
15	83	9.8	3354	4	US-10-160-758-11
16	83	9.8	3354	4	US-10-160-758-12
17	83	9.8	3354	4	US-10-174-677-9
18	83	9.8	3354	4	US-10-080-334-197
19	83	9.8	3354	4	US-10-080-334-198
20	83	9.8	3399	4	US-10-080-334-196
21	82.5	9.7	376	4	US-10-335-977-8864
22	82.5	9.7	376	4	US-10-335-977-8865
23	82.5	9.7	382	4	US-10-335-977-8866
24	82.5	9.7	420	4	US-10-437-963-180891
25	82	9.7	1638	5	US-10-732-923-8722
26	81	9.6	241	5	US-10-472-928-1270
27	81	9.6	783	4	US-10-282-122A-48649

28	80.5	9.5	2233	4	US-10-369-493-1998	Sequence 1998, Ap
29	79.5	9.4	367	4	US-10-425-115-294967	Sequence 294967, A
30	79.5	9.4	430	4	US-10-425-114-38846	Sequence 38846, A
31	79.5	9.4	934	5	US-10-840-512-228	Sequence 228, App
32	79.5	9.4	1452	4	US-10-282-122A-51753	Sequence 51753, A
33	79.5	9.4	1923	5	US-10-450-763-55533	Sequence 55533, A
34	79	9.3	662	5	US-10-732-923-8221	Sequence 8221, Ap
35	78	9.2	333	4	US-10-435-226-2	Sequence 2, Appli
36	78	9.2	628	4	US-10-389-566-1068	Sequence 1068, Ap
37	78	9.2	628	5	US-10-732-923-9638	Sequence 9638, Ap
38	78	9.2	1078	5	US-10-505-818-15	Sequence 15, Appl
39	77.5	9.1	206	5	US-10-732-923-8422	Sequence 8422, Ap
40	77	9.1	402	5	US-10-732-923-10596	Sequence 10596, A
41	77	9.1	402	5	US-10-732-923-10617	Sequence 10617, A
42	76.5	9.0	288	4	US-10-425-115-248274	Sequence 248274, A
43	76.5	9.0	406	5	US-10-501-282-1070	Sequence 1070, Ap
44	76.5	9.0	473	4	US-10-424-599-227005	Sequence 227005, A
45	76.5	9.0	594	4	US-10-425-115-248278	Sequence 248278, A

ALIGNMENTS

RESULT 1
US-09-808-602-95
; Sequence 95, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: No. US20020155115A1e1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 95
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-808-602-95

Query Match	10.1%	Score 85.5;	DB 3;	Length 464;
Best Local Similarity	23.1%	Pred. No. 3.9;		
Matches	30;	Conservative	26;	Mismatches 49; Indels 25; Gaps 5;
QY	9	GAGFTDQNKAYKEAMEALKENPTIDLENSYVPLDNOYK-----GIRVDEHPEYLHDKV	63	
Db	251	GDRMILTAHTVYPKDSVSLRKNQSVNVLGHTAIDEMKLGNHPVHRVWVHPDYRQNE-	309	
QY	64	WATATYNNDL-----NGIKTNDIMLGVIYIDPEEV-----GLGMELGYALSQOK	107	
Db	310	--SHNFSGDIALLELQHSIPLGNVLPVCLPDNETLYRSGLLGYVSGFGMGMGLTTEUK	367	
QY	108	YVLL-VIPDE	116	
Db	368	YSRLPVAPRE	377	

RESULT 2
US-09-808-602-93
; Sequence 93, Application US/09808602
; Patent No. US20020155115A1

```
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Heriman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: No. US20020155115A1el Proteins and Nuclec Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 93
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-808-602-93

Query Match      10.1%; Score 85.5; DB 3; Length 487;
Best Local Similarity 23.1%; Pred. No. 4.2;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;

QY      9 GAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNOYK-----GIRVDEHPEYLDKV 63
DB      274 GDRWILTAHTVYPKDSVSLRKNQSVNVFLGHTAIDEMLKGNHPVHRVVVHFDYRQNE- 332
QY      64 WATATYNNDL-----NGIKTNDIMLGVIYPDEEDV-----GLGMELGYALSQKG 107
DB      333 --SHNFGSDIALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGMGWLTTELK 390
QY      108 YVLL-VIPDE 116
DB      391 YSRLPVAPRE 400

RESULT 4
US-09-800-729-86
; Sequence 86, Application US/09800729
; Patent No. US20020068319A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 32 Human secreted proteins
; FILE REFERENCE: P2044P1
; CURRENT APPLICATION NUMBER: US/09/800,729
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: PCT/US00/26013
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 60/155,709
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 86
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-729-86

Query Match      10.0%; Score 84.5; DB 3; Length 396;
Best Local Similarity 23.1%; Pred. No. 4;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;

QY      9 GAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNOYK-----GIRVDEHPEYLDKV 63
DB      251 GDRWILTAHTVYPKDSVSLRKNQSVNVFLGHTAIDEMLKGNHPVHRVVVHFDYRQNE- 309
QY      64 WATATYNNDL-----NGIKTNDIMLGVIYPDEEDV-----GLGMELGYALSQKG 107
DB      310 --SHNFGSDIALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGMGWLTTELK 367
QY      108 YVLL-VIPDE 116
DB      368 YSRLPVAPRE 377

RESULT 5
US-09-833-245-2176
; Sequence 2176, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2176
; LENGTH: 396
; TYPE: PRT

; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Heriman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishna
; APPLICANT: Mezes, Peter S
; APPLICANT: Rastelli, Luca
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-697
; CURRENT APPLICATION NUMBER: US/09/800,198
; CURRENT FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 80
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-198-80

Query Match      10.1%; Score 85.5; DB 3; Length 487;
Best Local Similarity 23.1%; Pred. No. 4.2;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;
```

```
; ORGANISM: Homo sapiens
US-09-833-245-2176

Query Match      10.0%; Score 84.5; DB 3; Length 396;
Best Local Similarity 23.1%; Pred. No. 4;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;

Qy 9 GAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYK-----GIRVDEHPHYLHDKV 63
Db 251 GDRWILTAAHTIYPKDSVSLRKQSNVFLGHTAIDEMLKGNHPVHRVWVHPDYRQNE- 309
Qy 64 WATATYNDL-----NGIKTNDIMLGVIYPDEEDV-----GLGMELGYALSQK 107
Db 310 --SHNFGDIALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGMGWLTTELK 367
Qy 108 YVLL-VIPDE 116
Db 368 YSRLPVAPRE 377

RESULT 6
US-09-808-602-21
; Sequence 21, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: No. US20020155115A1el Proteins and Nuclec Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; PRIOR FILING DATE: 2001-03-14
; PRIOR FILING DATE: 2001-03-05
; PRIOR FILING DATE: 60/186,596
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 452
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-808-602-21

Query Match      10.0%; Score 84.5; DB 3; Length 452;
Best Local Similarity 23.1%; Pred. No. 4.8;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;

Qy 9 GAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYK-----GIRVDEHPHYLHDKV 63
Db 239 GDRWILTAAHTIYPKDSVSLRKQSNVFLGHTAIDEMLKGNHPVHRVWVHPDYRQNE- 297
Qy 64 WATATYNDL-----NGIKTNDIMLGVIYPDEEDV-----GLGMELGYALSQK 107
Db 298 --SHNFGDIALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGMGWLTTELK 355
Qy 108 YVLL-VIPDE 116
Db 356 YSRLPVAPRE 365

RESULT 7
US-09-808-602-17
; Sequence 17, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: No. US20020155115A1el Proteins and Nuclec Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; PRIOR FILING DATE: 2001-03-14
; PRIOR FILING DATE: 2001-03-05
; PRIOR FILING DATE: 60/186,596
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-808-602-17

Query Match      10.0%; Score 84.5; DB 3; Length 487;
Best Local Similarity 23.1%; Pred. No. 5.4;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;

Qy 9 GAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYK-----GIRVDEHPHYLHDKV 63
Db 274 GDRWILTAAHTIYPKDSVSLRKQSNVFLGHTAIDEMLKGNHPVHRVWVHPDYRQNE- 332
Qy 64 WATATYNDL-----NGIKTNDIMLGVIYPDEEDV-----GLGMELGYALSQK 107
Db 333 --SHNFGDIALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGMGWLTTELK 390
Qy 108 YVLL-VIPDE 116
Db 391 YSRLPVAPRE 400

RESULT 8
US-09-808-602-19
; Sequence 19, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: No. US20020155115A1el Proteins and Nuclec Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; PRIOR FILING DATE: 2001-03-14
; PRIOR FILING DATE: 2001-03-05
; PRIOR FILING DATE: 60/186,596
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-808-602-19

Query Match      10.0%; Score 84.5; DB 3; Length 487;
Best Local Similarity 23.1%; Pred. No. 5.4;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;

Qy 9 GAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYK-----GIRVDEHPHYLHDKV 63
```

```
Db      274 GDRWILTAHTIYPKDSVSRKQSVNVFLGHTAIDEMLKGNHPVHRVVHDPYRQNE- 332
Qy      64 WATATYNNDL-----NGIKTNDIMLGVIYPDEEDV-----GLGMELGYALSQK 107
Db      333 --SHNFGSDIALLELQHSIPLGPNVLPVCLPDNETLYRSLGLGVSVSGFGMEMGLTTTELK 390
Qy      108 YVLL-VIPDE 116
Db      391 YSRLPVAPRE 400

RESULT 9
US-09-800-198-17
; Sequence 17, Application US/09800198
; Publication No. US20030087816A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Cornie AM
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrmann, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishna
; APPLICANT: Mezes, Peter S
; APPLICANT: Rastelli, Luca
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-697
; CURRENT APPLICATION NUMBER: US/09/800,198
; CURRENT FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-198-17
```

```
Query Match      10.0%; Score 84.5; DB 3; Length 487;
Best Local Similarity 23.1%; Pred. No. 5.4;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;

Qy      9 GAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYK-----GIRVDEHPEYLHDKV 63
Db      274 GDRWILTAHTIYPKDSVSRKQSVNVFLGHTAIDEMLKGNHPVHRVVHDPYRQNE- 332
Qy      64 WATATYNNDL-----NGIKTNDIMLGVIYPDEEDV-----GLGMELGYALSQK 107
Db      333 --SHNFGSDIALLELQHSIPLGPNVLPVCLPDNETLYRSLGLGVSVSGFGMEMGLTTTELK 390
Qy      108 YVLL-VIPDE 116
Db      391 YSRLPVAPRE 400
```

```
RESULT 10
US-09-800-198-19
; Sequence 19, Application US/09800198
; Publication No. US20030087816A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Cornie AM
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrmann, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishna
; APPLICANT: Mezes, Peter S
; APPLICANT: Rastelli, Luca
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-697
; CURRENT APPLICATION NUMBER: US/09/800,198
; CURRENT FILING DATE: 2001-03-05
```

```
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-198-19

Query Match      10.0%; Score 84.5; DB 3; Length 487;
Best Local Similarity 23.1%; Pred. No. 5.4;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;

Qy      9 GAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYK-----GIRVDEHPEYLHDKV 63
Db      274 GDRWILTAHTIYPKDSVSRKQSVNVFLGHTAIDEMLKGNHPVHRVVHDPYRQNE- 332
Qy      64 WATATYNNDL-----NGIKTNDIMLGVIYPDEEDV-----GLGMELGYALSQK 107
Db      333 --SHNFGSDIALLELQHSIPLGPNVLPVCLPDNETLYRSLGLGVSVSGFGMEMGLTTTELK 390
Qy      108 YVLL-VIPDE 116
Db      391 YSRLPVAPRE 400

RESULT 11
US-11-097-143-27759
; Sequence 27759, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27759
; LENGTH: 1537
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-27759
```

```
Query Match      9.9%; Score 84; DB 6; Length 1537;
Best Local Similarity 20.4%; Pred. No. 30;
Matches 42; Conservative 30; Mismatches 48; Indels 86; Gaps 11;

Qy      9 GAGWFTDRQ-----NKAYKEAMEALKENPTI-----DLENSYVPLDN 45
Db      584 GWDWIDDEEDSCGSDNDHDKPKVE---EQTATEDATDKAQTGNDDEKADLITKAKVEDD 640
Qy      46 QYKGIRVDEHPEY-----LHDKV-----WATATYNNDLNGIKTN 79
```

```

641 Ey---RTEQTYYSIAHTIHERKVBQASIMVNGTLKEYQIKGLEWLSYNNNLINGILAD 697
Db      | : | : | | | : | : | : | : | : | : | : | : | : | : | : | : |
80 DIMLG-----VVIPDEEDVGLGMELGVALSGKGKVLIV----LPD-----EDYGKP 121
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
698 EMGLGTIQTISLVLYLMDRKV-----MGPLYIIPLSLTPNWLVSFEKKAPA 746
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
122 INLMSWGSDNVIKM--SQLKDFNF 145
Qy      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
747 VGVVSYKSGPOGRRLQNOMRATKEN 772
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 12

```

US-10-732-923-8723
; Sequence 8723, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8723
; LENGTH: 1634
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-732-923-8723

```

Query Match	9.9%;	Score 84;	DB 5;	Length 1634;	
Best Local Similarity	20.4%;	Pred. No. 32;			
Matches	42;	Conservative 30;	Mismatches 48;	Indels 86;	Gaps 11;
Qy	9	GAGWFTORQ-----NKAYKEAMEALKENPTI-----	DLENSYVPLDN	45	
Db	681	GWDWIDDEEDSGSNDHDHKPVE-----EQPTATEDATDKAQTGDNDEDAKDLITRAKVEDD	737		
Qy	46	QYKGIRVDEHPY-----LHDKV-----WATATYNDLNGIKTN	79		
Db	738	EY---RTEBQTYYSIAHTIHEKVKVEQASIMVNGTILKEYQIRGLEWLVSLYNNLNGILAD	794		
Qy	80	DIMLG-----VYIPDESDVGLGMELGVALSGOKVLLV-----IPD-----EDYCKP	121		
Db	795	EMGLGKTTIQTISLVTYIMDRKV-----MGPLYIIVPLSTLEPNWVLEFEKWAPA	843		
Qy	122	INLMSWGVSDNVIRM--SQLKDFNFN	145		
Db	844	GVVSYKSGSPQGRRLLOMQMRATKFN	869		

RESULT 13

```

RES001.13
US-10-732-923-8724
; Sequence 8724, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8724
; LENGTH: 1638
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-732-923-8724

```

Query Match 9.9%; Score 84; DB 5; Length 1638;
Best Local Similarity 20.4%; Pred. No. 32;

Matches	42;	Conservative	30;	Mismatches	48;	Indels	86;	Gaps	11;
Qy	9	GAGFTDRQ	-----NKAYKEAMALKENPTI	-----	-----	-----	-----	-----	-----
Db	685	GWDWIDDEEDSCGSDNDHKPKVE	---EQPTATEDATDKAQTGDNDKAKVDEDD	----	----	----	----	----	----
Qy	46	OYKGRVDEHPEY	-----LHDKV	-----	-----	-----	-----	-----	-----
Db	742	EY---RTBEQTYYSIAHTHEKVVQASIMVNGTLKEYQYKGLWVLSLYNNNNLNGILAD	-----	-----	-----	-----	-----	-----	-----
Qy	80	DIMLG	-----VYIPDEEDVGLGELGYALSQGYVLLV	-----	-----	-----	-----	-----	-----
Db	799	EMGLGKTIQIISLVYLMDRKKV	-----	-----	-----	-----	-----	-----	-----
Qy	122	INLMSWGSVDNVIKQ	-----SQLKDFNFN	-----	-----	-----	-----	-----	-----
Db	848	GVVVSYKGSPOGRLLQNOMRYKFN	-----	-----	-----	-----	-----	-----	-----

RESULT 14

```

US-11-097-143-12630
; Sequence 12630, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 12630
; LENGTH: 1638
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-12630

```

Query Match	9.9%	Score 84	DB 6	Length 1638
Best Local Similarity	20.4%	Pred. No. 32		
Matches 42	Conservative 30	Mismatches 48	Indels 86	Gaps 11
Qy	9	GAGFTTDRQ-----NKAYKEAMEALKENPTI-----	DLENSVYPLDN	45
Db	685	GWDWIDDEEDSCGSDNDHDKPVE---EQPTATEDATDKAQTGDNDEDAKDLITRAKAKVEDD	741	
Qy	46	QYKGIRVDEHPEY-----LHDKV-----WATATYNDLNGIKTN	79	
Db	742	EY---RTEQYYSIAHTIHEKWEQASIMVNGTLKEYQIKGLEWLVSLYNNLNGILAD	798	
Qy	80	DIMLG-----VYIPDEEDVGLGMELGYALSGQKYVLV---	IPD-----EDYCKP	121
Db	799	EMGLGKTTQTSIVTYLMDRKKV-----MGPYLIIVPLSTLPNNWVLEFEKWA	847	
Qy	122	INLMSWGVSDNWIKM---SOLKDFNFN	145	

```

Db      848   VGVVSYKSGPQGRLLQNQMRAKFN    873
          :::||: | :: |:|:: ||
RESULT 15
US-10-160-758-11
; Sequence 11, Application US/10160758
; Publication No. US20030036076A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: CADS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-089C
; CURRENT APPLICATION NUMBER: US/10/160,758
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 60/296,076
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,605
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/357,253
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 11
; LENGTH: 3354
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-160-758-11

Query Match           9.8%; Score 83; DB 4; Length 3354;
Best Local Similarity 22.2%; Pred.No. 1.1e+02;
Matches 36; Conservative 32; Mismatches 68; Indels 26; Gaps 8;

QY       15 DRONKAYKEAMEALKE--NPTIDL-----ENSVPLDNO-YKGIRVDEHPEYLHD 61
        ||: ::: : :: ||: ||: |: : : |:: |:
DB     1593 DRRQSFYHLVAIVADEGDTPTSATTHVVITVDENDNAFMFQQPHYEVILLDEGDFTLNT 1652
                -LNGLTKTNDIMLGVTIPD---BEDVGL---GMELGYALSOGKYVLAVI 113
QY       62 KWMTATAYNND--LNGLTKTNDIMLGVTIPD---BEDVGL---GMELGYALSOGKYVLAVI 113
SLITIQAQLDLDEGGNGIVTTVAIVAGNIWTFRIDRHMGVITAAKELYEISHGRYTTLTVT 1712
DB     1653 SLITIQAQLDLDEGGNGIVTTVAIVAGNIWTFRIDRHMGVITAAKELYEISHGRYTTLTVT 1712

PY     114 PDSDYGKPINLMMSGVDSNVIKMSQDKDFPNPKPRFD-FYBG 154
              :         :             :
DB     1713 ATDO----CPILSHRLSTTVLVNNVANDINDNVPTFFPRDYEG 1750
              :         :             :
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2005, 22:05:00 ; Search time 12 Seconds

(without alignments)
93.321 Million cell updates/sec

Title: US-10-049-750-14

Perfect score: 848

Sequence: 1 MPKKTIFYGAGWFTDRQNK.....QLKDFNFKPRDFYEGAVY 157

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 54001 seqs, 7132810 residues

Total number of hits satisfying chosen parameters: 54001

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	78	9.2	US-11-070-627-4	Sequence 4, Appli
2	73.5	8.7	US-10-793-626-1948	Sequence 1948, Ap
3	73	8.6	US-11-074-176-202	Sequence 202, App
4	70.5	8.3	US-11-070-627-3	Sequence 3, Appli
5	69.5	8.2	US-10-979-821-12	Sequence 12, Appl
6	69.5	8.2	US-11-070-627-2	Sequence 2, Appli
7	69.5	8.2	US-11-070-627-9	Sequence 9, Appli
8	68.5	8.1	US-10-793-626-668	Sequence 668, App
9	68.5	8.1	US-10-485-517-145	Sequence 145, App
10	68	8.0	US-11-094-586-4	Sequence 4, Appli
11	68	8.0	US-10-821-234-1662	Sequence 1662, Ap
12	68	8.0	US-11-070-627-1	Sequence 1, Appli
13	68	8.0	US-10-821-234-1007	Sequence 1007, Ap
14	67.5	8.0	US-11-004-057-6	Sequence 6, Appli
15	67.5	8.0	US-11-004-057-21	Sequence 21, Appl
16	67	7.9	US-10-630-203-2	Sequence 2, Appli
17	67	7.9	US-11-103-037-1	Sequence 1, Appli
18	67	7.9	US-10-392-234A-14	Sequence 14, Appl
19	66.5	7.8	US-10-467-962B-4	Sequence 4, Appli
20	66.5	7.8	US-10-858-730-71	Sequence 71, Appl
21	66.5	7.8	US-11-091-643-4	Sequence 4, Appli
22	66	7.8	US-11-070-627-8	Sequence 8, Appli
23	66	7.8	US-11-070-627-10	Sequence 10, Appl
24	66	7.8	US-11-017-550-65	Sequence 65, Appl
25	65.5	7.7	US-10-873-528-144	Sequence 144, App

ALIGNMENTS

RESULT 1

US-11-070-627-4
; Sequence 4, Application US/11070627
; Publication No. US20050271625A1
; GENERAL INFORMATION:
; APPLICANT: Nash, Kevin R.
; APPLICANT: Burger, Corinna
; TITLE OF INVENTION: TAAV-NEPRILYSIN COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 36689.8
; CURRENT APPLICATION NUMBER: US/11/070.627
; CURRENT FILING DATE: 2005-03-02
; PRIOR APPLICATION NUMBER: 60/549,399
; PRIOR FILING DATE: 2004-03-02
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 4
; LENGTH: 750
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-11-070-627-4

Query Match	9.2%	Score 78;	DB 7;	Length 750;
Best Local Similarity	23.4%	Pred. No. 3.4;		
Matches	44;	Conservative	25;	Mismatches 41;
				Indels 78;
				Gaps 11;
QY	1	MPKKTIFYGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYKQIRVDEH-----	55	
Db	236	LPSRDYBCTG-----IYKEACTAY-----VDFMLAVAKLIQREGLPIDENQISVE	272	
QY	56	-----PEYLHDKWATATYNN-----DLNGIK-----TNDIM- 82		
Db	273	MNKVMEKEIKANATTKSEDRNDPMLLYNKMTLAQIQNFSLEINKGPFWSNFTNEIMS	332	
QY	83	LGVIYPIDEEDVGLGMEIGYALSQGYVLLVTPDEDYGPPI-----NLSWG-VS	130	
Db	333	TWNINIFNEEDV-----VYVA---PEYLIKU-----KPILTKYFPRDFQNLFSWRFIM	377	
QY	131	DNVIMKMSQ	138	
Db	378	DLVSSLSR	385	

RESULT 2

US-10-793-626-1948
; Sequence 1948, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN

	Matches	32;	Conservative	16;	Mismatches	52;	Indels	29;	Gaps	6;
Qy	22	KEAMEAL	KNPTID	---	LENSYV	PLDNQY	KGIRV	-----	DEHP	SYLHDKVWATATYN 70
							:	:	:	
Db	157	KCLEETIKENPTIKAVILNYP	TNPTGVEYSEDEIKALAKV	IKDNHLYVITDEIYSTLY	---	---	---	---	---	---
Qy	71	NDLNGIKTNDIM	----	LGVI	YP--	DEEDVGLG	MELGYAL	SGQKYVLLVITPDEIDYGRKIN 123		
Db	216	----	GVKHFSIASLI	PERA	IVISGL	SKSHAMTGYRL	GVVAGPAKIM	-----	AEIGKVHG 265	
Qy	124	LMSWGVS	SDN 132							
Db	266	LMVTTT	IDS 274							

```

RESULT 6
US-11-070-627-2
; Sequence 2, Application US/11070627
; Publication No. US20050271625A1
; GENERAL INFORMATION:
; APPLICANT: Nash , Kevin R.
; APPLICANT: Burger, Corinna
; TITLE OF INVENTION: RAAV-NEPRILYSIN COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 36689.8
; CURRENT APPLICATION NUMBER: US/11/070,627
; CURRENT FILING DATE: 2005-03-02
; PRIOR APPLICATION NUMBER: 60/549,399
; PRIOR FILING DATE: 2004-03-02
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 750
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-11-070-627-2

```

Query Match	8.2%;	Score 69.5;	DB 7;	Length 750;
Best Local Similarity	23.5%;	Pred. NO. 23;		
Matches	47;	Conservative 29;	Mismatches 55;	Indels 69; Gaps 12;
Qy	1	MPKKTIVFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLNQYKIGRVDEH-----	55	
Db	226	LPSRDYECTG-----IYKEACTAY-----VDFMISVARLFEQRLPIDENQLSLE	272	
Qy	56	-----PEYLHDKWATATYNN-----DLNGIK-----TNDIM-	82	
Db	273	MNKVMELEKEIATATTKPEDRNDPMLLYNKYMTLAKLQNNFSLIEINGKCFWSNFTNEIMS	332	
Qy	83	-LGWYIPDEEDVGLGMEIGYA-----LSQCKYVLLVLPDEDYCKPINLMSWG-VSDNVIKWS	137	
Db	333	TWNINIQHEEV-----VYIAPYITLKPILTKYSPDLQ-----NLSWRFTMDLVSSLS	384	
Qy	138	QLKDFNFNKPDPFVEGAVY	157	
Db	385	R-----NYKESR-NAPRKALY	399	

```

RESULT 7
US-11-070-627-9
; Sequence 9, Application US/11070627
; Publication No. US20050271625A1
; GENERAL INFORMATION:
; APPLICANT: Nash , Kevin R.
; APPLICANT: Burger, Corinna
; TITLE OF INVENTION: rAAV-NEPRILYSIN COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 36689.8
; CURRENT APPLICATION NUMBER: US/11/070,627
; CURRENT FILING DATE: 2005-03-02
; PRIOR APPLICATION NUMBER: 60/549,399
; PRIOR FILING DATE: 2004-03-02
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 9

```

```

; LENGTH: 786
; TYPE: PRT
; ORGANISM: Canis familiaris
US-11-070-627-9

Query Match      8.2% Score 69.5; DB 7; Length 786;
Best Local Similarity 24.9%; Pred. No. 24;
Matches 42; Conservative 28; Mismatches 58; Indels 41; Gaps 11;

Qy      9 GAGWFTDRONKAYKEAMEALKENPTTIDLENSYV-----PLDNQYKGRIVDEHPHPEYLHKV 63
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      288 GNGLIRKEKGLLIDENQLSLEMRVMELEKEIASATTKPEDR-----NDPMLLYNKM 339

Qy      64 WATATYNN-----DLNGTK-----TNDIM--LGVIPEEDVDVGLMELGYA---LSQSKYV 109
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      340 TLAQIQNNFTLEIDGGPFPSWNSFTNEMTVMTNININEEEV-----VYAYPEYLTUKLI 394

Qy      110 LLVIPEDYDGKPINLMSWG-VSDNVIMKMSQLKDFNFKPRDFYEGAVY 157
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      395 LTKYSSRDLO---NLMSWRIFIMDLVSSLR-----NYKESR-NAPFRKALY 435
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 8
US-10-793-626-668
; Sequence 668, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 668
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-668

```

Query Match	8.1%;	Score 68.5;	DB 6;	Length 395;
Best Local Similarity	18.1%;	Pred. No. 12;		
Matches	41;	Conservative 33;	Mismatches 58;	Indels 95; Gaps 10.

QY	14	TDRN-KAYKEAMEALK-----ENPTIDLENS-----YVPLDNQYKGI RVDH	55
Db	179	TPRNKEITYENATESIPKOLNDPKQTASVTKPTLEYQSTNEMYAEVYNQLKPIRET	238
QY	56	PEY-----LHDKVWATATYN-----	70
Db	239	QGYNWCAGYTMSALFNATYNTNRYNAESVMRYLHPNLRGHDFOFTGLTSMELRFGRSQ	298
QY	71	-----NDLNGIKTND--TMLGVYIIPDEEDVGLGMELGYA-----LSOCKVYL	110
Db	299	GRNTQYLNRMSTSYNVEDQLTNNOGIAVLGRKRVESDGIHACHAMAVAGNAGNVNGQKVI	358
QY	111	LVI PDEYDGKFINLMSGVSDNVIKMSQLKDPFNFKPRDFYEGAVY	157
Db	359	LINWPDPNG-----LMTODAHENIIPVS-----NGDHYEWY-ASIIY	393

RESULT 9
US-10-485-517-145
; Sequence 145, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated

QY	114	-PDEYKPKINLMGVSNDVNIKMSQKDFNFNKRDFDYEGAVY	157
Db	230	DPVSDY---LELM-----ENVDFQLIKDL-LSRPDRFFIDAMH	265
RESULT 11			
US-10-821-234-1662			
Sequence 1662, Application US/10821234			
Publication No. US20050255114A1			
GENERAL INFORMATION:			
APPLICANT: Labat, Ivan			
APPLICANT: Stache-Crain, Birgit			
APPLICANT: Andarmani, Susan			
APPLICANT: Tang, Y. Tom			
TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia			
FILE REFERENCE: 821A			
CURRENT APPLICATION NUMBER: US/10/821,234			
CURRENT FILING DATE: 2004-04-07			
PRIOR APPLICATION NUMBER: US 60/462,047			
PRIOR FILING DATE: 2003-04-07			
NUMBER OF SEQ ID NOS: 1704			
SOFTWARE: pt_seq_genes Version 1.0			
SEQ ID NO 1662			
LENGTH: 747			
TYPE: PRT			
ORGANISM: Homo sapiens			
US-10-821-234-1662			
Query Match	8.0%;	Score 68;	DB 6; Length 747;
Best Local Similarity	38.2%;	Pred. No. 32;	
Matches	13;	Conservative	7; Mismatches 12; Indels 2; Gaps 1;
QY	114	PDEYKPKINLMGVSNDVNIKMSQKDFNFNKP	147
Db	612	PDEIQRPVRVPSWGLDNVVCSPAR--NFSRP	643
RESULT 12			
US-11-070-627-1			
Sequence 1, Application US/11070627			
Publication No. US20050271625A1			
GENERAL INFORMATION:			
APPLICANT: Nash, Kevin R.			
APPLICANT: Burger, Corinna			
TITLE OF INVENTION: xAAV-NEPRILYSIN COMPOSITIONS AND METHODS OF USE			
FILE REFERENCE: 36889.8			
CURRENT APPLICATION NUMBER: US/11/070,627			
CURRENT FILING DATE: 2005-03-02			
PRIOR APPLICATION NUMBER: 60/549,399			
PRIOR FILING DATE: 2004-03-02			
NUMBER OF SEQ ID NOS: 11			
SOFTWARE: PatentIn version 3.3			
SEQ ID NO 1			
LENGTH: 750			
TYPE: PRT			
ORGANISM: Homo sapiens			
US-11-070-627-1			
Query Match	8.0%;	Score 68;	DB 7; Length 750;
Best Local Similarity	24.3%;	Pred. No. 32;	
Matches	44;	Conservative	23; Mismatches 50; Indels 64; Gaps 10;
QY	1	MPKTIYFCAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNOYKGRVDEH	55
Db	226	LPSRDYECTG-----IYREACTAY-----VDFMISVARLIQBERLPIDENQALAE	272
QY	56	-----PEYLHDKVWATATYNN---DLNGIK-----TNDIM-	82
Db	273	MNKVMELEKEIANATAKPEDRNDPMLLYNKNRQLAQIQNFSLEINGKPPFSLNFTNEIMS	332
QY	83	-LGVYIPDEEDVGLGMELGYA---LSQKXVLLVVPDEYKPKINLMGVSNDVNIKMSQKDFNFNKRDFDYEGAVY	137
Db	333	TWNISITNEEDV-----VVAPEYVLTIKLPILTYKSARDLQ---NLMMSWRFIMDLVSSLS	384

Qy 138 Q 138
Db 385 R 385

RESULT 13

US-10-821-234-1007
; Sequence 1007, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt SEQ_genes Version 1.0
; SEQ ID NO 1007
; LENGTH: 751
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1007

Query Match 8.0%; Score 68; DB 6; Length 751;
Best Local Similarity 24.3%; Pred. No. 32;
Matches 44; Conservative 23; Mismatches 50; Indels 64; Gaps 10;
Qy 1 MPKTIYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYKQIRVDEH-----55
Db 227 LPSRDYECTG-----IYKEACTAY-----VDFMISVARLIRQEBRLPIDENQALE 273
Qy 56 -----PEYLDKQWATATYNN-----DLNGIK-----TNDIM- 82
Db 274 MNKVMLEKEIANATAKPEDRNDPMLYNNKRLAQIQQNFSLEINGKPFSLNFTNIMS 333
Qy 83 -LGVIYVDEEDVGLGMEIGYA--L SQGYVLLVDPEDYDKPKNLMSWG-VSDNVIKMS 137
Db 334 TVNISITNEEDV-----VVIYAPEYLTUKPIITKYSARDLQ---NLMSWRFIMDLVSSLS 385
Qy 138 Q 138
Db 386 R 386

RESULT 14

US-11-004-057-6
; Sequence 6, Application US/11004057
; Publication No. US20050244846A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Gary L.
; TITLE OF INVENTION: MEK1 PROTEINS AND FRAGMENTS THEREOF FOR USE IN REGULATING
; FILE REFERENCE: CPI-042CPPC
; CURRENT APPLICATION NUMBER: US/11/004,057
; CURRENT FILING DATE: 2004-12-02
; PRIOR APPLICATION NUMBER: US/09/403,075
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1302
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-004-057-6

Query Match 8.0%; Score 67.5; DB 7; Length 1302;
Best Local Similarity 23.9%; Pred. No. 74;

Matches 37; Conservative 21; Mismatches 62; Indels 35; Gaps 6;
Qy 18 NKAYKEAMEALKEN-----PTIDLENSYVPLDNOYKQIRVDEHPEYL 59
Db 964 NQCKEKMEAEAEALAIAMAMSASQDALPIVPOQVENGEDII-----IIQDTPETL 1017
Qy 60 HDKVMATATYNNDLNGIKTNDIMLGVIYIP--DEEDVGLGMEIGYALSQGYVLLVDPED 117
Db 1018 PGHTKAKQPYREDAEWLKGQOIGLGAFFSSCYQADVGTTGLM--AVKQVTVRNTSSQEQ 1075
Qy 118 -----YGKPINLMSWGVSDNVIKM--SQLKDFNFN 145
Db 1076 EVVEALREERIMMSHLNHPNIIRMLGATCEKSNYN 1110

RESULT 15

US-11-004-057-21
; Sequence 21, Application US/11004057
; Publication No. US20050244846A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Gary L.
; TITLE OF INVENTION: MEK1 PROTEINS AND FRAGMENTS THEREOF FOR USE IN REGULATING
; FILE REFERENCE: CPI-042CPPC
; CURRENT APPLICATION NUMBER: US/11/004,057
; CURRENT FILING DATE: 2004-12-02
; PRIOR APPLICATION NUMBER: US/09/403,075
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 1493
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-11-004-057-21

Query Match 8.0%; Score 67.5; DB 7; Length 1493;
Best Local Similarity 24.5%; Pred. No. 89;
Matches 38; Conservative 20; Mismatches 62; Indels 35; Gaps 7;
Qy 18 NKAYKEAMEALKEN-----PTIDLENSYVPLDNOYKQIRVDEHPEYL 59
Db 1155 NQCKEKMEAEAEALAIAMAMSASQDALPIVPOQVENG-----EDIIIIQADT-PETL 1208
Qy 60 HDKVMATATYNNDLNGIKTNDIMLGVIYIP--DEEDVGLGMEIGYALSQGYVLLVDPED 117
Db 1209 PGHTKANEPYREDTEWLKQOIGLGAFFSSCYQADVGTTGLM--AVKQVTVRNTSSQEQ 1266
Qy 118 -----YGKPINLMSWGVSDNVIKM--SQLKDFNFN 145
Db 1267 EVVEALREERIMMSHLNHPNIIRMLGATCEKSNYN 1301

Search completed: December 23, 2005, 22:20:50
Job time : 19 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2005, 21:57:09 ; Search time 569 Seconds
(without alignments)
381.313 Million cell updates/sec

Title: US-10-049-750-14
Perfect score: 848
Sequence: 1 MPKTYIFGAGWFTDRQNA.....QLKDFNFKNRPFDFYEGAVY 157

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 7861189 seqs, 1381955077 residues
Total number of hits satisfying chosen parameters: 7861189

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA Main:*

1: /cgn2_6/ptodata/1/paa/pctus COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US066 COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US073 COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US074 COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US075 COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US076 COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US077 COMB.pep:*
8: /cgn2_6/ptodata/1/paa/US078 COMB.pep:*
9: /cgn2_6/ptodata/1/paa/US079 COMB.pep:*
10: /cgn2_6/ptodata/1/paa/US080 COMB.pep:*
11: /cgn2_6/ptodata/1/paa/US081 COMB.pep:*
12: /cgn2_6/ptodata/1/paa/US082 COMB.pep:*
13: /cgn2_6/ptodata/1/paa/US083 COMB.pep:*
14: /cgn2_6/ptodata/1/paa/US084 COMB.pep:*
15: /cgn2_6/ptodata/1/paa/US085 COMB.pep:*
16: /cgn2_6/ptodata/1/paa/US086 COMB.pep:*
17: /cgn2_6/ptodata/1/paa/US087 COMB.pep:*
18: /cgn2_6/ptodata/1/paa/US088 COMB.pep:*
19: /cgn2_6/ptodata/1/paa/US089 COMB.pep:*
20: /cgn2_6/ptodata/1/paa/US090 COMB.pep:*
21: /cgn2_6/ptodata/1/paa/US091 COMB.pep:*
22: /cgn2_6/ptodata/1/paa/US092 COMB.pep:*
23: /cgn2_6/ptodata/1/paa/US093 COMB.pep:*
24: /cgn2_6/ptodata/1/paa/US094 COMB.pep:*
25: /cgn2_6/ptodata/1/paa/US095 COMB.pep:*
26: /cgn2_6/ptodata/1/paa/US096 COMB.pep:*
27: /cgn2_6/ptodata/1/paa/US097 COMB.pep:*
28: /cgn2_6/ptodata/1/paa/US098 COMB.pep:*
29: /cgn2_6/ptodata/1/paa/US099 COMB.pep:*
30: /cgn2_6/ptodata/1/paa/US100 COMB.pep:*
31: /cgn2_6/ptodata/1/paa/US101 COMB.pep:*
32: /cgn2_6/ptodata/1/paa/US102 COMB.pep:*
33: /cgn2_6/ptodata/1/paa/US103 COMB.pep:*
34: /cgn2_6/ptodata/1/paa/US104 COMB.pep:*
35: /cgn2_6/ptodata/1/paa/US105 COMB.pep:*
36: /cgn2_6/ptodata/1/paa/US106 COMB.pep:*
37: /cgn2_6/ptodata/1/paa/US107 COMB.pep:*
38: /cgn2_6/ptodata/1/paa/US108 COMB.pep:*
39: /cgn2_6/ptodata/1/paa/US109 COMB.pep:*
40: /cgn2_6/ptodata/1/paa/US110 COMB.pep:*
41: /cgn2_6/ptodata/1/paa/US111 COMB.pep:*
42: /cgn2_6/ptodata/1/paa/US112 COMB.pep:*
43: /cgn2_6/ptodata/1/paa/US114 COMB.pep:*

44: /cgn2_6/ptodata/1/paa/US600 COMB.pep:*
45: /cgn2_6/ptodata/1/paa/US601 COMB.pep:*
46: /cgn2_6/ptodata/1/paa/US602 COMB.pep:*
47: /cgn2_6/ptodata/1/paa/US603 COMB.pep:*
48: /cgn2_6/ptodata/1/paa/US604 COMB.pep:*
49: /cgn2_6/ptodata/1/paa/US605 COMB.pep:*
50: /cgn2_6/ptodata/1/paa/US606 COMB.pep:*
51: /cgn2_6/ptodata/1/paa/US607 COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	848	100.0	157	30	US-10-049-750-14	Sequence 14, Appl
2	728	85.8	158	40	US-11-097-292-2	Sequence 2, Appl
3	728	85.8	158	40	US-11-097-292-14	Sequence 14, Appl
4	719	84.8	159	40	US-11-097-292-12	Sequence 12, Appl
5	719	84.8	159	50	US-60-622-712-264	Sequence 264, App
6	719	84.8	159	51	US-60-711-491-264	Sequence 264, App
7	629	74.2	133	40	US-11-097-292-10	Sequence 10, Appl
8	620	73.1	165	27	US-09-791-537-8009	Sequence 8009, Ap
9	405	47.8	84	40	US-11-097-292-8	Sequence 8, Appl
10	209	24.6	168	40	US-11-097-292-6	Sequence 6, Appl
11	166.5	19.6	167	40	US-11-097-292-4	Sequence 4, Appl
12	154.5	18.2	167	50	US-60-622-712-2174	Sequence 2174, Ap
13	154.5	18.2	167	51	US-60-711-491-2174	Sequence 2174, Ap
14	91.5	10.8	156	34	US-10-417-884A-6487	Sequence 6487, Ap
15	91.5	10.8	156	34	US-10-417-884A-6487	Sequence 6487, Ap
16	91	10.7	719	27	US-09-791-537-10210	Sequence 10210, A
17	91	10.7	719	27	US-09-791-537-96101	Sequence 96101, A
18	85.5	10.1	438	26	US-09-629-469A-18384	Sequence 18384, A
19	85.5	10.1	438	39	US-10-917-503-18384	Sequence 18384, A
20	85.5	10.1	456	34	US-10-455-772-22	Sequence 22, Appl
21	85.5	10.1	464	1	PCT-US00-04340-48	Sequence 48, Appl
22	85.5	10.1	464	25	US-09-507-209-48	Sequence 48, Appl
23	85.5	10.1	464	28	US-09-808-602-95	Sequence 95, Appl
24	85.5	10.1	487	28	US-09-800-198-80	Sequence 80, Appl
25	85.5	10.1	487	28	US-09-808-602-93	Sequence 93, Appl
26	85.5	10.1	487	34	US-10-455-772-16	Sequence 16, Appl
27	85.5	10.1	487	34	US-10-455-772-46	Sequence 46, Appl
28	85.5	10.1	487	34	US-10-455-772-48	Sequence 48, Appl
29	85.5	10.1	487	34	US-10-455-772-50	Sequence 50, Appl
30	85.5	10.1	487	34	US-10-455-772-52	Sequence 52, Appl
31	85.5	10.1	487	38	US-10-844-295-80	Sequence 80, Appl
32	84.5	10.0	198	45	US-60-173-469-1175	Sequence 1175, Ap
33	84.5	10.0	198	45	US-60-198-955-1449	Sequence 1449, Ap
34	84.5	10.0	243	34	US-10-455-772-38	Sequence 38, Appl
35	84.5	10.0	247	34	US-10-455-772-20	Sequence 20, Appl
36	84.5	10.0	396	1	PCT-US01-11988-2176	Sequence 2176, Ap
37	84.5	10.0	396	28	US-09-833-245-2176	Sequence 2176, Ap
38	84.5	10.0	396	28	US-09-833-245A-2176	Sequence 2176, Ap
39	84.5	10.0	396	28	US-09-833-245B-2176	Sequence 86, Appl
40	84.5	10.0	396	34	US-10-458-714-66	Sequence 21, Appl
41	84.5	10.0	452	28	US-09-808-602-21	Sequence 42, Appl
42	84.5	10.0	452	34	US-10-455-772-42	Sequence 42, Appl
43	84.5	10.0	456	34	US-10-455-772-18	Sequence 18, Appl
44	84.5	10.0	456	34	US-10-455-772-24	Sequence 24, Appl
45	84.5	10.0	456	34	US-10-455-772-30	Sequence 30, Appl

ALIGNMENTS

RESULT 1
US-10-049-750-14
; Sequence 14, Application US/10049750
; GENERAL INFORMATION:
; APPLICANT: Tischer, Wilhelm

```

; APPLICANT: Ihlenfeldt, Hans-Georg
; APPLICANT: Barzu, Octavian
; APPLICANT: Sakamoto, Hiroshi
; APPLICANT: Pistotnik, Elisabeth
; APPLICANT: Marliere, Philippe
; APPLICANT: Pochet, Sylvie
; TITLE OF INVENTION: Enzymatic synthesis of deoxyribonucleosides
; FILE REFERENCE: 20373PWO Deoxyribonucleosides
; CURRENT APPLICATION NUMBER: US/10/049,750
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: BP99116425.2
; PRIOR FILING DATE: 1999-08-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 14
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Lactobacillus leichmannii
US-10-049-750-14

Query Match      100.0%; Score 848; DB 30; Length 157;
Best Local Similarity 100.0%; Pred. No. 7.3e-87;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPKTTYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSVPLDNQYKGRVDEHPEYLH 60
Db 1 MPKTTYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSVPLDNQYKGRVDEHPEYLH 60

Qy 61 DKWATATYNDLNGIKTNDIMLGVIYIPDEEDVGLGMEALYALSGQKYLVLVLPDEDYDK 120
Db 61 DKWATATYNDLNGIKTNDIMLGVIYIPDEEDVGLGMEALYALSGQKYLVLVLPDEDYDK 120

Qy 121 PINLMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157
Db 121 PINLMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157

RESULT 2
US-11-097-292-2
; Sequence 2, Application US/11097292
; GENERAL INFORMATION:
; APPLICANT: KAMINSKI, PIERRE-ALEXANDRE
; APPLICANT: TAILLIEZ, PATRICK
; APPLICANT: MARLIERE, PHILIPPE
; APPLICANT: QUENEZ, PASCAL
; APPLICANT: COTAYA, RACHEL
; TITLE OF INVENTION: LACTOBACILLUS N-DEOXYRIBOSYL TRANSFERASES, CORRESPONDING
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES AND THEIR USES
; FILE REFERENCE: 250246USOXPT
; CURRENT APPLICATION NUMBER: US/11/097,292
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: FR 0 111 911
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/FR02/03120
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Lactobacillus helveticus NTD
US-11-097-292-2

Query Match      85.8%; Score 728; DB 40; Length 158;
Best Local Similarity 82.6%; Pred. No. 2.8e-73;
Matches 128; Conservative 20; Mismatches 7; Indels 0; Gaps 0;

Qy 3 KTTYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSVPLDNQYKGRVDEHPEYLHDK 62
Db 4 KKTLYFGAGWFEKQNKAYKEAMEALKENPTVDLENSVPLENQYKGRIDEHPEYLHNI 63

Qy 63 VWAATATYNDLNGIKTNDIMLGVIYIPDEEDVGLGMEALYALSGQKYLVLVLPDEDYDKPI 122
Db 64 EWASATYHNDLVGIKTSDVNLGVYLPDEEDVGLGMEALYALSGQKYLVLVLPDEDYDKPI 123

Qy 123 NLMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157
Db 124 NLMSGVCDNAIKISELKDFDNKPRYFNFDGAVY 158

RESULT 4
US-11-097-292-12
; Sequence 12, Application US/11097292
; GENERAL INFORMATION:
; APPLICANT: KAMINSKI, PIERRE-ALEXANDRE
; APPLICANT: TAILLIEZ, PATRICK
; APPLICANT: MARLIERE, PHILIPPE
; APPLICANT: QUENEZ, PASCAL
; APPLICANT: COTAYA, RACHEL
; TITLE OF INVENTION: LACTOBACILLUS N-DEOXYRIBOSYL TRANSFERASES, CORRESPONDING
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES AND THEIR USES
; FILE REFERENCE: 250246USOXPT
; CURRENT APPLICATION NUMBER: US/11/097,292
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: FR 0 111 911
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/FR02/03120
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 12
; LENGTH: 159
; TYPE: PRT

```

```
; ORGANISM: Lactobacillus acidophilus NTD
US-11-097-292-12

Query Match      84.8%; Score 719; DB 40; Length 159;
Best Local Similarity 83.1%; Pred. No. 2.9e-72;
Matches 128; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

QY 4 KTIYFGAGWFTDRQNKAYKAAMEALKENPTIDLENSYVPLDNOYKGRVDEHPEYVLDKV 63
Db 6 KTLYFGAGWFTDRQNKAYKAAMEALKONPTVDLENSYVPLENQYKDIRVDEHPEYVLDIE 65

QY 64 WATATYNDLNGIKTNDIMLGVIIPDEEDVGLGMELGVALSGQKYVLLVIPDEDYGKPIN 123
Db 66 WASATYHNDLIGIKSSDMLGVLPEDVDVGLGMELGVALSGQKYVLLVIPDEDYGKPIN 125

QY 124 LMSWGVSDNVVKMSQLKDFNFKPRDFYEGAVY 157
Db 126 LMSWGVCDNAIKISELKDFDNFKPRFNYDGAVY 159

RESULT 5
US-60-622-712-264
; Sequence 264, Application US/60622712
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd
; APPLICANT: Russell, William
; APPLICANT: Alterman, Eric
; TITLE OF INVENTION: LACTOBACILLUS ACIDOPHILUS NUCLEIC ACIDS AND USES THEREOF
; FILE REFERENCE: 5051-604PR10
; CURRENT APPLICATION NUMBER: US/60/622,712
; CURRENT FILING DATE: 2004-10-27
; NUMBER OF SEQ ID NOS: 2559
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 264
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
US-60-622-712-264

Query Match      84.8%; Score 719; DB 50; Length 159;
Best Local Similarity 83.1%; Pred. No. 2.9e-72;
Matches 128; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

QY 4 KTIYFGAGWFTDRQNKAYKAAMEALKENPTIDLENSYVPLDNOYKGRVDEHPEYVLDKV 63
Db 6 KTLYFGAGWFTDRQNKAYKAAMEALKONPTVDLENSYVPLENQYKDIRVDEHPEYVLDIE 65

QY 64 WATATYNDLNGIKTNDIMLGVIIPDEEDVGLGMELGVALSGQKYVLLVIPDEDYGKPIN 123
Db 66 WASATYHNDLIGIKSSDMLGVLPEDVDVGLGMELGVALSGQKYVLLVIPDEDYGKPIN 125

QY 124 LMSWGVSDNVVKMSQLKDFNFKPRDFYEGAVY 157
Db 126 LMSWGVCDNAIKISELKDFDNFKPRFNYDGAVY 159

RESULT 6
US-60-711-491-264
; Sequence 264, Application US/60711491
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd
; APPLICANT: Russell, William
; APPLICANT: Alterman, Eric
; TITLE OF INVENTION: LACTOBACILLUS ACIDOPHILUS NUCLEIC ACIDS AND USES THEREOF
; FILE REFERENCE: 035051/296336
; CURRENT APPLICATION NUMBER: US/60/711,491
; CURRENT FILING DATE: 2005-08-26
; NUMBER OF SEQ ID NOS: 2561
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 264
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
```

```
US-60-711-491-264

Query Match      84.8%; Score 719; DB 51; Length 159;
Best Local Similarity 83.1%; Pred. No. 2.9e-72;
Matches 128; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

QY 4 KTIYFGAGWFTDRQNKAYKAAMEALKENPTIDLENSYVPLDNOYKGRVDEHPEYVLDKV 63
Db 6 KTLYFGAGWFTDRQNKAYKAAMEALKONPTVDLENSYVPLENQYKDIRVDEHPEYVLDIE 65

QY 64 WATATYNDLNGIKTNDIMLGVIIPDEEDVGLGMELGVALSGQKYVLLVIPDEDYGKPIN 123
Db 66 WASATYHNDLIGIKSSDMLGVLPEDVDVGLGMELGVALSGQKYVLLVIPDEDYGKPIN 125

QY 124 LMSWGVSDNVVKMSQLKDFNFKPRDFYEGAVY 157
Db 126 LMSWGVCDNAIKISELKDFDNFKPRFNYDGAVY 159

RESULT 7
US-11-097-292-10
; Sequence 10, Application US/11097292
; GENERAL INFORMATION:
; APPLICANT: KAMINSKI, PIERRE-ALEXANDRE
; APPLICANT: TAILLIEZ, PATRICK
; APPLICANT: MARLIERE, PHILIPPE
; APPLICANT: QUENEZ, PASCAL
; APPLICANT: COTAYA, RACHEL
; TITLE OF INVENTION: LACTOBACILLUS N-DEOXYRIBOSYL TRANSFERASES, CORRESPONDING
; FILE REFERENCE: 250246US0XPC7
; CURRENT APPLICATION NUMBER: US/11/097,292
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: FR 0 111 911
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/FR02/03120
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 10
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Lactobacillus amylovorus NTD
US-11-097-292-10

Query Match      74.2%; Score 629; DB 40; Length 133;
Best Local Similarity 84.2%; Pred. No. 3.5e-62;
Matches 112; Conservative 17; Mismatches 4; Indels 0; Gaps 0;

QY 25 MEALKENPTIDLENSYVPLDNOYKGRVDEHPEYVLDKVWATATYNDLNGIKTNDIMLG 84
Db 1 MEALKKNPTVDLENSYVPLDNOYKGRVDEHPEYVLDIEWASSTYHNDLVGKSSDMLG 60

QY 85 VYIPDEEDVGLGMELGVALSGQKYVLLVIPDEDYGKPINLMSWGVSDNVVKMSQLKDFNF 144
Db 61 VYLPEDVDVGLGMELGVALSGQKYVLLVIPDEDYGKPINLMSWGVCDNVKISELKDFDF 120

QY 145 NKPRDFYEGAVY 157
Db 121 NRPRFNYDGAVY 133

RESULT 8
US-09-791-537-8009
; Sequence 8009, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
```



```
RESULT 15
US-10-417-884A-6487
; Sequence 6487, Application US/10417884A
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10417,884A
; FILING DATE: 17-Apr-2003
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6487:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...156
; SEQUENCE DESCRIPTION: SEQ ID NO: 6487:
US-10-417-884A-6487

Query Match      10.8%; Score 91.5; DB 34; Length 156;
Best Local Similarity 23.0%; Pred. No. 0.28;
Matches 35; Conservative 28; Mismatches 74; Indels 15; Gaps 5;

Qy      6 IYFGAGHFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNQYKGIKRVDEHPEYLRDKVWA 65
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      14 IYLAGPFFSEQIDRVSRIEKALEENKTV--TSFYSPRHQ-----ESNYELFSAGWA 64

Qy      66 TATYNNDLNGIKTNDIMLGVIYIPDEE--DVGLMGELGVALSQGYVLLVIPDEDYGRPIN 123
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      65 QEVYEKOMEELTNAEFVVAILDPEHQITDPGTAYELGVATMLKPKMIIV---QETVTPTN 121

Qy      124 LMSGVSDNVIKMSQ-LKDFNFKPRDFYEG 154
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      122 LMITQSLHTYIKSQQAVREYDFETLPVETVYG 153
```

Search completed: December 23, 2005, 22:15:22
Job time : 572 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2005, 22:00:24 ; Search time 14 Seconds
(without alignments)
118.825 Million cell updates/sec

Title: US-10-049-750-14

Perfect score: 848

Sequence: 1 MPKTIYFGAGWFTDRQNK.....QLKDFNFKRFDYEGAVY 157

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 62187 seqs, 10595856 residues

Total number of hits satisfying chosen parameters: 62187

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA New:*

- 1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/2/paa/US11_NEW_COMB.pep.*
- 8: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84.5	10.0	396	7	US-11-264-096-2176
2	83	9.8	3354	8	US-60-732-162-1678
3	74	8.7	357	7	US-11-045-004-1742
4	73	8.6	1310	6	US-10-556-060-326
5	72.5	8.5	578	8	US-60-732-162-404
6	72.5	8.5	578	8	US-60-742-219-464
7	72.5	8.5	710	7	US-11-288-493-44
8	72	8.5	893	6	US-10-244-081A-28
9	70.5	8.3	844	7	US-11-127-877A-40
10	70.5	8.3	844	8	US-60-732-162-402
11	70.5	8.3	844	8	US-60-742-219-462
12	70.5	8.3	899	8	US-60-732-162-408
13	70.5	8.3	899	8	US-60-742-219-466
14	70.5	8.3	961	8	US-60-732-162-406
15	70.5	8.3	961	8	US-60-742-219-460
16	70.5	8.3	1576	7	US-11-238-031-5
17	69.5	8.2	192	7	US-11-045-004-607
18	69	8.1	138	7	US-11-045-004-1152
19	69	8.1	743	1	PCT-US05-20516-2
20	69	8.1	749	8	US-60-742-871-285
21	69	8.1	750	8	US-60-742-871-287
22	69	8.1	750	8	US-60-742-871-288
23	69	8.1	750	8	US-60-742-871-289
24	69	8.1	750	8	US-60-742-871-290
25	69	8.1	750	8	US-60-742-871-291

26	69	8.1	750	8	US-60-742-871-292
27	69	8.1	782	1	PCT-US05-20516-4
28	69	8.1	831	8	US-60-742-219-884
29	69	8.1	831	8	US-60-742-871-509
30	69	8.1	831	8	US-60-742-871-511
31	69	8.1	838	8	US-60-742-871-510
32	68	8.0	750	8	US-60-742-871-286
33	67	7.9	721	7	US-11-170-482-12
34	67	7.9	738	7	US-11-045-004-1782
35	66.5	7.8	283	7	US-11-045-004-2575
36	66.5	7.8	580	7	US-11-045-004-938
37	66.5	7.8	779	8	US-60-733-588-11
38	66.5	7.8	779	8	US-60-733-434-11
39	66.5	7.8	902	8	US-60-733-588-9
40	66.5	7.8	902	8	US-60-733-434-9
41	66.5	7.8	1220	1	PCT-US05-11978-6
42	66	7.7	1196	8	US-60-732-162-784
43	65.5	7.7	248	8	US-60-741-048-42
44	65.5	7.7	466	8	US-60-742-872-59
45	65.5	7.7	466	8	US-60-742-872-60

ALIGNMENTS

RESULT 1

US-11-264-096-2176
; Sequence 2176, Application US/11264096
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PFS46D1
; CURRENT APPLICATION NUMBER: US/11/264,096
; CURRENT FILING DATE: 2005-11-02
; PRIOR APPLICATION NUMBER: 09/833,245
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229, 358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256, 931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199, 384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2176
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-264-096-2176

Query Match 10.0%; Score 84.5; DB 7; Length 396;
Best Local Similarity 23.1%; Pred. NO. 0.072;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;

QY	9	GAGWTTDRQNKAYKEAMEALKENPTIDLENSVPLDNOVK-----GIRVDEHPYLVHDKV 63
DB	251	GDWILTTAAHTTTPKDSVSLRKNQSVNVLGHTALDEMLKLGNNHPVHVHPDTRQNS- 309
QY	64	WATATYNNDL-----NGIKTNDIMLGVIIPDEEDV-----GLGMELGYALSQK 107
DB	310	--SHNFSGDIALLELOHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFCMGWMLTTEUK 367
QY	108	VYLL-VIPDE 116
DB	368	YSRLPVAPRE 377

RESULT 2

US-60-732-162-1678
; Sequence 1678, Application US/60732162
; GENERAL INFORMATION:
; APPLICANT: Belouchi, Abdelmajid
; APPLICANT: Raelson, John V

```
; APPLICANT: Bradley, Walter E
; APPLICANT: Paquin, Bruno
; APPLICANT: Fournier, Helene
; APPLICANT: Nguyen-Huu, Quynh
; APPLICANT: Croteau, Pascal
; APPLICANT: Allard, Rene
; APPLICANT: Debrus, Sophie
; APPLICANT: Berdewegh, Paul V
; APPLICANT: Little, Randall D
; APPLICANT: Keith, Tim
; APPLICANT: Segal, Jonathan
; TITLE OF INVENTION: Genemap of the Human Genes Associated With Asthma Disease
; FILE REFERENCE: 059908-5010-PR
; CURRENT APPLICATION NUMBER: US/60/732,162
; CURRENT FILING DATE: 2005-11-02
; NUMBER OF SEQ ID NOS: 417
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1678
; LENGTH: 3354
; TYPE: PRT
; ORGANISM: Homosapiens
US-60-732-162-1678

Query Match      9.8%; Score 83; DB 8; Length 3354;
Best Local Similarity 22.2%; Pred. No. 1.9;
Matches 36; Conservative 32; Mismatches 68; Indels 26; Gaps 8;

Qy 15 DRQKVKKEAMEALKE--NPTIDL-----ENSVPVLDNQ-YKGIRVDEHPEYVLDH 61
Db 1593 DRKQSFHLVAIVDEGTPTLSATTVYTVIVDENDNAPMFQQPHYEVLLDEGDFTLNT 1652

Qy 62 KVMATATYNNND--LNGIKTNDIMLGVIYPD---EEDVGL---GMELGYALSQGGKYVLLVI 113
Db 1653 SLITIQALDLDEGNGTVTVVAIVAGNVNTFRIDRHGVITAAKELQYVESHGRTYTLIVT 1712

Qy 114 PDEDYGRPINLMSGVSDNVIKMSQLKDFNFKRPD-FYEG 154
Db 1713 ATDQ----CPILSHRLTSTTVLVNVVNDINDNVTFPRDYEG 1750

RESULT 3
US-11-045-004-1742
; Sequence 1742, Application US/11045004
; GENERAL INFORMATION:
; APPLICANT: BUCHRIESER, CARMEN
; APPLICANT: FRANGUL, LIONEL
; APPLICANT: COUVE, ELISABETH
; APPLICANT: RUSNOK, CHRISTOPHE
; APPLICANT: FSIHI, HAFIDA
; APPLICANT: DEHOUX, PIERRE
; APPLICANT: DUSSURGET, OLIVIER
; APPLICANT: CHETOJANI, FARID
; APPLICANT: NEDJARI, HAFED
; APPLICANT: GLASER, PHILIPPE
; APPLICANT: KUNST, FRANCK
; APPLICANT: COSSART, PASCALE
; APPLICANT: DANIELS, JUSTIN
; APPLICANT: GOEBEL, WERNER
; APPLICANT: KREFT, JURGEN
; APPLICANT: KUHN, MICHAEL
; APPLICANT: NG, EVA
; APPLICANT: VAZQUEZ-BOLAND, ANTONIO
; APPLICANT: DOMINGUEZ-BERNAL, GUSTAVO
; APPLICANT: GARRIDO-GARCIA, PATRICIA
; APPLICANT: TIERREZ-MARTINEZ, ALBERTO
; APPLICANT: AMEND, ALEXANDRA
; APPLICANT: CHAKRABORTY, TRINAD
; APPLICANT: DOMANN, EUGEN
; APPLICANT: HAIN, THORSTEN
; APPLICANT: BERCHE, PATRICK
; APPLICANT: CHARBIT, ALAIN
; APPLICANT: DURANT, LIONEL
; APPLICANT: PEREZ-DIAZ, JOSE-CLAUDIO
```

```
; APPLICANT: BAQUERO, FERNANDO
; APPLICANT: GARCIA DEL PORTILLO, FRANCISCO
; APPLICANT: GOMEZ-LOPEZ, NURIA
; APPLICANT: MADUENIO, ENCARNA
; APPLICANT: PABLOS, BETRIZ DE
; APPLICANT: WEHLAND, JURGEN
; APPLICANT: KARST, UWE
; APPLICANT: ENTIAN, KARL-DIETER
; APPLICANT: HAUP, JORG
; APPLICANT: ROSE, MATTHIAS
; APPLICANT: VOSS, HAMUT
; TITLE OF INVENTION: LISTERIA MONOCYTOGENES GENOME, POLYPEPTIDES AND USES
; FILE REFERENCE: 05394.0018-02
; CURRENT APPLICATION NUMBER: US/11/045,004
; CURRENT FILING DATE: 2005-01-28
; PRIOR APPLICATION NUMBER: 10/637,657
; PRIOR FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: 10/257,023
; PRIOR FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: PCT/FR01/01118
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: FR 00/04,629
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 2854
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1742
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Listeria monocytogenes
US-11-045-004-1742

Query Match      8.7%; Score 74; DB 7; Length 357;
Best Local Similarity 26.6%; Pred. No. 0.91;
Matches 25; Conservative 18; Mismatches 33; Indels 18; Gaps 4;

Qy 29 KENPTDLENSVYP-----LDNQYKGRVDEHPEYLDHKKWAT--ATYNNDLNGIKTNDIM 82
Db 99 EENLLPLDHSKLPNEKYLDPRFMDLSFDDDNKYSMPYFWGTGLIYINKEMFPDKNFDTW 158

Qy 83 LGVYIPD-----EEDVGLGM-ELGYALS 104
Db 159 NALFDPFLKNQIILLIDGAREVWGLNSLGYSLN 192

RESULT 4
US-10-556-060-326
; Sequence 326, Application US/10556060
; GENERAL INFORMATION:
; APPLICANT: MEINKE, ANDREAS
; APPLICANT: NAGY, ESZTER
; APPLICANT: HANNER, MARKUS
; APPLICANT: HORKY, MARKUS
; APPLICANT: KALLEDA, SABINE
; APPLICANT: PRUSTOMERSKY, SONJA
; TITLE OF INVENTION: S. AGALACTIAE ANTIGENS I + II
; FILE REFERENCE: SONN:080US
; CURRENT APPLICATION NUMBER: US/10/556,060
; CURRENT FILING DATE: 2005-11-07
; PRIOR APPLICATION NUMBER: PCT/EP2004/004856
; PRIOR FILING DATE: 2004-05-06
; PRIOR APPLICATION NUMBER: 03450112.2
; PRIOR FILING DATE: 2003-05-07
; PRIOR APPLICATION NUMBER: 03450266.6
; PRIOR FILING DATE: 2003-11-28
; NUMBER OF SEQ ID NOS: 514
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 326
; LENGTH: 1310
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-10-556-060-326

Query Match      8.6%; Score 73; DB 6; Length 1310;
```

[illegible]

RESULT 5

```

US-60-732-162-404
; Sequence 404, Application US/60732162
; GENERAL INFORMATION:
; APPLICANT: Abdelmajid
; APPLICANT: Belouchi, John V
; APPLICANT: Raelson, John V
; APPLICANT: Bradley, Walter E
; APPLICANT: Paquin, Bruno
; APPLICANT: Fournier, Helene
; APPLICANT: Nguyen-Huu, Quynh
; APPLICANT: Croteau, Pascal
; APPLICANT: Allard, Rene
; APPLICANT: Debrus, Sophie
; APPLICANT: Berdewegh, Paul V
; APPLICANT: Little, Randall D
; APPLICANT: Keith, Tim
; APPLICANT: Segal, Jonathan
; TITLE OF INVENTION: Genemap of the Human Genes Associated with Asthma Disease
; FILE REFERENCES: 059908-5010-PR
; CURRENT APPLICATION NUMBER: US/60732,162
; CURRENT FILING DATE: 2005-11-02
; NUMBER OF SEQ ID NOS: 4417
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 404
; LENGTH: 578
; TYPE: PRT
; ORGANISM: Homosapiens
; US-60-732-162-404

```

Query Match	8.5%;	Score 72.5;	DB 8;	Length 578;
Best Local Similarity	23.9%;	Pred. No. 2.6;		
Matches	39;	Conservative 20;	Mismatches 51;	Indels 53; Gaps 8;
Qy	3	KTITYFGAGWFTDRQNKAY-----KEAMEALKENPTIDL-----	ENSYV	41
Db	386	KKYWFLIGWYADWNFKIYDPSINCTVDMEATEAVEGHITTEIVMLNPANTRSISNMTSQE	445	
Qy	42	PLDNQYKGIIRVDEHPE-----YLHKVWATATYNNDLGKIKTINDIMLGVIPODEE	92	
Db	446	FVEKLTK--RLKRPEETGGQEAFLAYDAFWALALN-----KTS-----	485	
Qy	93	VGLGMEIYALSOQKYVLLVTPDEYDGKPINLMW-GVSDNVI	134	
Db	486	-GGSGRSGVRLEDFTYNNOTTITDQIY-RAMNSSFEGVSGHV	526	

9 T. J. S. J. A.

US-60-742-219-464
SECTION 6
SEQUENCE 464, Application US/60742219
GENERAL INFORMATION:
APPLICANT: Belouchi, Abdelmajid
APPLICANT: Raelson, John Verner
APPLICANT: Bradley, Walter Edward
APPLICANT: Paquin, Bruno
APPLICANT: Nguyen-Huu, Quynh
APPLICANT: Croteau, Pascal
APPLICANT: Allard, Rene
APPLICANT: Little, Randall David
APPLICANT: Keith, Tim
APPLICANT: Cousineau, Johanne
APPLICANT: Eerdewegh, Paul Van

```

; APPLICANT: Segal, Jonathan
; TITLE OF INVENTION: Gene Map of the Human Genes Associated with Psoriasis
; FILE REFERENCE: 306522-2000
; CURRENT APPLICATION NUMBER: US/60/742,219
; CURRENT FILING DATE: 2005-12-05
; NUMBER OF SEQ ID NOS: 7303
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 464
; LENGTH: 578
; TYPE: prt
; ORGANISM: Homo sapiens
; US-60-742-219-464

```

Query Match	8.5%;	Score 72.5;	DB 8;	Length 578;
Best Local Similarity	23.9%;	Pred. No. 2.6;		
Matches 39;	Conservative 20;	Mismatches 51;	Indels 53;	Gaps 8;
Qy	3	KTIIYFGAGFTDRQNKAY-----KEAMEALKENPTIDL-----	ENSYV	41
Db	386	KCVWFLLIGIYADNFWKIYDPSINCTVDTEAVEGHITTEIVMLNPANTRISNMTSQE		445
Qy	42	PLDNQYKGIIRVDBEHP-----YLHDKVWATATYNNDLNGIKTNDIMLGVIYIPDEED		92
Db	446	FVEKLTK--RLKRHPETGGFQEQAPLAYDAIWALALALN-----KTS-----		485
Qy	93	VGLGMEIGLYALSOCKYVLLVLPIDEDYKCPINLMSW--GVSDNVI		134
Db	486	-CGGSRSGVLEDFNYNNQITTDIYI--RAMNSSFEGVSGHVV		526

RESULT 7

```

US-11-288-493-44
;
; Sequence 44, Application US/11288493
; GENERAL INFORMATION:
; APPLICANT: Kelly, Louise M.
; APPLICANT: Carroll, Joseph M.
; APPLICANT: Farlow, Deborah
; APPLICANT: Healy, Aileen
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING
; TITLE OF INVENTION: HEMATOLOGICAL DISORDERS USING 9118, 990, 17662, 81982, 630,
; TITLE OF INVENTION: 21472, 17692, 19290, 21620, 21689, 28899, 53659, 64549,
; TITLE OF INVENTION: 9465, 23544, 7366, 27417, 57259, 21844, 943, 2061, 5891,
; TITLE OF INVENTION: 9137, 13908, 14310, 17600, 25584, 27824, 28469, 38947,
; TITLE OF INVENTION: 53003, 965, 56639, 9661, 16052, 1521, 6662, 13913, 12405 OR
; TITLE OF INVENTION: 5014
; FILE REFERENCE: MPI03-015P1RNMNM1
; CURRENT APPLICATION NUMBER: US/11/288,493
; CURRENT FILING DATE: 2005-11-29
; PRIOR APPLICATION NUMBER: US/10/772,636
; PRIOR FILING DATE: 2004-02-05
; PRIOR APPLICATION NUMBER: US 60/445,241
; PRIOR FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US 60/448,389
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/456,320
; PRIOR FILING DATE: 2003-03-20
; PRIOR APPLICATION NUMBER: US 60/460,279
; PRIOR FILING DATE: 2003-04-03
; PRIOR APPLICATION NUMBER: US 60/465,924
; PRIOR FILING DATE: 2003-04-28
; PRIOR APPLICATION NUMBER: US 60/470,052
; PRIOR FILING DATE: 2003-05-13
; PRIOR APPLICATION NUMBER: US 60/498,106
; PRIOR FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: US 60/500,179
; PRIOR FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: US 60/502,909
; PRIOR FILING DATE: 2003-09-15
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 710

```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-288-493-44

Query Match      8.5%; Score 72.5; DB 7; Length 710;
Best Local Similarity 26.0%; Pred. No. 3.4;
Matches 46; Conservative 26; Mismatches 54; Indels 51; Gaps 13;

Qy 9 GAGWFTDRQNKAYKEAMEALKENPTIDLE-----NSYVPLDNOYKGI 50
Db 147 GSDVVTIKPMK-----VDGAKELPDV-LERVKFSCMAWTHDGKGMFYNSYPQQDGKSDGT 200
Qy 51 RVDEHPYLDHKWATATYNNDLNGIKTNDIMLGVIYPDEEDVGLGMEIGYALSGQKYVL 110
Db 201 ETSTN---LHQKL-----YHVLGTQSEDILCAEF-PDEPKWVGGAELS---DDGRYVL 248
Qy 111 LVTPDEYDGKPINLM-----SWGVDNVIKMSQLKDFENKPRDFY--EGAVY 157
Db 249 LSI--REGCDPVNRLWYCDLQESSGIA-GILKWVKLID-NF-EGEYDVTNGETVF 300

RESULT 8
US-10-244-081A-28
; Sequence 28, Application US/10244081A
; GENERAL INFORMATION:
; APPLICANT: Gerard, Gary
; APPLICANT: Shandilya, Harini
; APPLICANT: Griffiths, Katherine Rachel
; APPLICANT: Gibbs, Moreland David
; APPLICANT: Leonard, Peter
; APPLICANT: Bergquist, Peter Leonard
; APPLICANT: Potter, Jason
; TITLE OF INVENTION: DNA Polymerases and Mutants Thereof
; FILE REFERENCE: 0942.536001
; CURRENT APPLICATION NUMBER: US/10/244,081A
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 60/318,903
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3641
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Thermatoga neopolitina
US-10-244-081A-28

Query Match      8.5%; Score 72; DB 6; Length 893;
Best Local Similarity 24.1%; Pred. No. 5.3;
Matches 39; Conservative 29; Mismatches 66; Indels 28; Gaps 8;

Qy 11 GWFTDRQNKAYKEAMEALKENPT--IDLENSYV-PLDNOYKGI RVDEHPY-----LHDK 62
Db 297 GYEIVKHQTFEDIEKLEVPSPALDLETSSLPFNCIEVIGISVSKPKYAYIPLHR 356
Qy 63 VWATATYNNDLNGI--KTNDIMLGVIYPDEEDVGLGMEIGYALSGQKYVLVIPD----- 115
Db 357 ----NAHMLDRTLVLKLEILED---PSSKIVGQNLKYDKVLMWKGISVPYHPFDTWI 409
Qy 116 -----EDYGPKNLMGVSNDVNIKMSQLKDFENKPRDFP 151
Db 410 AAYLLEPNKKFNLEDLSLFLGYKMTSYQELMSFSPLFGF 451

RESULT 9
US-11-127-877A-40
; Sequence 40, Application US/11127877A
; GENERAL INFORMATION:
; APPLICANT: Merckhiers, Pascal G.
; APPLICANT: Hoffmann, Marcel
; APPLICANT: Spittaels, Koenraad F. F.
; APPLICANT: Laenen, Wendy
; TITLE OF INVENTION: Methods, Compositions and Compound Assays For Inhibiting
```

```
; TITLE OF INVENTION: Amyloid-Beta Protein Production
; FILE REFERENCE: P27,800-B USA
; CURRENT APPLICATION NUMBER: US/11/127,877A
; CURRENT FILING DATE: 2005-05-12
; PRIOR APPLICATION NUMBER: 60/570,352
; PRIOR FILING DATE: 2004-05-12
; PRIOR APPLICATION NUMBER: 60/603,948
; PRIOR FILING DATE: 2004-08-24
; NUMBER OF SEQ ID NOS: 590
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 40
; LENGTH: 844
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-127-877A-40

Query Match      8.3%; Score 70.5; DB 7; Length 844;
Best Local Similarity 23.9%; Pred. No. 7.1;
Matches 39; Conservative 20; Mismatches 51; Indels 53; Gaps 8;

Qy 3 KKTIFYGAGWFTDRQNKAY-----KEAMEALKENPTIDL-----ENSYV 41
Db 269 KKYVFLIGWYADNWFKIYDPSINCTVDTEAVEGHITTEIVMLNPANTRISINMTSQE 328
Qy 42 PLDNOYKGI RVDEHP-----YLHDKWATATYNNDLNGIKTNDIMLGVIYPDEED 92
Db 329 FVEKLTK--RLKRHPBETGGFOEAPLAYDAIWAIALALN-----KTS----- 368
Qy 93 VGLGMELGYALSGQKYVLVIPDEEDYKCPINLMMSW-GVSDNVI 134
Db 369 -GGGRSGVRLEDFNYYNNQTITDQIY-RAMNSSFEGVSGHV 409

RESULT 10
US-60-732-162-402
; Sequence 402, Application US/60732162
; GENERAL INFORMATION:
; APPLICANT: Belouchi, Abdelmajid
; APPLICANT: Raelson, John V
; APPLICANT: Bradley, Walter E
; APPLICANT: Paquin, Bruno
; APPLICANT: Fournier, Helene
; APPLICANT: Nguyen-Huu, Quynh
; APPLICANT: Croteau, Pascal
; APPLICANT: Allard, Rene
; APPLICANT: Debrus, Sophie
; APPLICANT: Berdewegh, Paul V
; APPLICANT: Little, Randall D
; APPLICANT: Keith, Tim
; APPLICANT: Segal, Jonathan
; TITLE OF INVENTION: Genemap of the Human Genes Associated With Asthma Disease
; FILE REFERENCE: 059908-5010-PR
; CURRENT APPLICATION NUMBER: US/60/732,162
; CURRENT FILING DATE: 2005-11-02
; NUMBER OF SEQ ID NOS: 4417
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 402
; LENGTH: 844
; TYPE: PRT
; ORGANISM: Homosapiens
US-60-732-162-402

Query Match      8.3%; Score 70.5; DB 8; Length 844;
Best Local Similarity 23.9%; Pred. No. 7.1;
Matches 39; Conservative 20; Mismatches 51; Indels 53; Gaps 8;

Qy 3 KKTIFYGAGWFTDRQNKAY-----KEAMEALKENPTIDL-----ENSYV 41
Db 269 KKYVFLIGWYADNWFKIYDPSINCTVDTEAVEGHITTEIVMLNPANTRISINMTSQE 328
Qy 42 PLDNOYKGI RVDEHP-----YLHDKWATATYNNDLNGIKTNDIMLGVIYPDEED 92
Db 329 FVEKLTK--RLKRHPBETGGFOEAPLAYDAIWAIALALN-----KTS----- 368
```

Qy 93 VGLGMELGYALSQGYVLLVLPDEDYDKPILMSW-GVSDNVI 134
Db 369 -GGGSGVRLDFNYYNQTTDQIY-RAMNSSSFEVSGHV 409

RESULT 11

US-60-742-219-462
; Sequence 462, Application US/60742219
; GENERAL INFORMATION:
; APPLICANT: Belouchi, Abdelmajid
; APPLICANT: Raelson, John Verner
; APPLICANT: Bradley, Walter Edward
; APPLICANT: Paquin, Bruno
; APPLICANT: Nguyen-Huu, Quynh
; APPLICANT: Croteau, Pascal
; APPLICANT: Allard, Rene
; APPLICANT: Little, Randall David
; APPLICANT: Keith, Tim
; APPLICANT: Cousineau, Johanne
; APPLICANT: Eerdewegh, Paul Van
; APPLICANT: Segal, Jonathan
; TITLE OF INVENTION: Gene Map of the Human Genes Associated with Psoriasis
; FILE REFERENCE: 306522-2000
; CURRENT APPLICATION NUMBER: US/60/742,219
; CURRENT FILING DATE: 2005-12-05
; NUMBER OF SEQ ID NOS: 7303
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 462
; LENGTH: 844
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-742-219-462

Query Match 8.3%; Score 70.5; DB 8; Length 844;
Best Local Similarity 23.9%; Pred. No. 7.1; Indels 53; Gaps 8;
Matches 39; Conservative 20; Mismatches 51; Indels 53; Gaps 8;
Qy 3 KKTIFYGAGWFTDRQNKAY-----KEAMEALKENPTIDL-----ENSYV 41
Db 269 KKYVWFLIGWADNWFKYDPSINCTVDEMTAEVGHITTEIVMLNPANTRISISNMTSQE 328
Qy 42 PLDNOYKQIRVDEHPE-----YLDKQWATATYNDLNGIKTNDIMLGVIYPDEED 92
Db 329 FVEKLTGK--RLKRHPETGGFQEAFLAYDAIWAALALN-----KTS----- 368
Qy 93 VGLGMELGYALSQGYVLLVLPDEDYDKPILMSW-GVSDNVI 134
Db 369 -GGGSGVRLDFNYYNQTTDQIY-RAMNSSSFEVSGHV 409

RESULT 12

US-60-732-162-408
; Sequence 408, Application US/60732162
; GENERAL INFORMATION:
; APPLICANT: Belouchi, Abdelmajid
; APPLICANT: Raelson, John V
; APPLICANT: Bradley, Walter E
; APPLICANT: Paquin, Bruno
; APPLICANT: Fournier, Helene
; APPLICANT: Nguyen-Huu, Quynh
; APPLICANT: Croteau, Pascal
; APPLICANT: Allard, Rene
; APPLICANT: Debrus, Sophie
; APPLICANT: Eerdewegh, Paul V
; APPLICANT: Little, Randall D
; APPLICANT: Keith, Tim
; APPLICANT: Segal, Jonathan
; TITLE OF INVENTION: Genemap of the Human Genes Associated With Asthma Disease
; FILE REFERENCE: 059908-5010-PR
; CURRENT APPLICATION NUMBER: US/60/732,162
; CURRENT FILING DATE: 2005-11-02
; NUMBER OF SEQ ID NOS: 4417

; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 408
; LENGTH: 899
; TYPE: PRT
; ORGANISM: Homosapiens
US-60-732-162-408

Query Match 8.3%; Score 70.5; DB 8; Length 899;
Best Local Similarity 23.9%; Pred. No. 7.8; Indels 53; Gaps 8;
Matches 39; Conservative 20; Mismatches 51; Indels 53; Gaps 8;
Qy 3 KKTIFYGAGWFTDRQNKAY-----KEAMEALKENPTIDL-----ENSYV 41
Db 324 KKYVWFLIGWADNWFKYDPSINCTVDEMTAEVGHITTEIVMLNPANTRISISNMTSQE 383
Qy 42 PLDNOYKQIRVDEHPE-----YLDKQWATATYNDLNGIKTNDIMLGVIYPDEED 92
Db 384 FVEKLTGK--RLKRHPETGGFQEAFLAYDAIWAALALN-----KTS----- 423
Qy 93 VGLGMELGYALSQGYVLLVLPDEDYDKPILMSW-GVSDNVI 134
Db 424 -GGGSGVRLDFNYYNQTTDQIY-RAMNSSSFEVSGHV 464

RESULT 13

US-60-742-219-466
; Sequence 466, Application US/60742219
; GENERAL INFORMATION:
; APPLICANT: Belouchi, Abdelmajid
; APPLICANT: Raelson, John Verner
; APPLICANT: Bradley, Walter Edward
; APPLICANT: Paquin, Bruno
; APPLICANT: Nguyen-Huu, Quynh
; APPLICANT: Croteau, Pascal
; APPLICANT: Allard, Rene
; APPLICANT: Little, Randall David
; APPLICANT: Keith, Tim
; APPLICANT: Cousineau, Johanne
; APPLICANT: Eerdewegh, Paul Van
; APPLICANT: Segal, Jonathan
; TITLE OF INVENTION: Gene Map of the Human Genes Associated with Psoriasis
; FILE REFERENCE: 306522-2000
; CURRENT APPLICATION NUMBER: US/60/742,219
; CURRENT FILING DATE: 2005-12-05
; NUMBER OF SEQ ID NOS: 7303
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 466
; LENGTH: 899
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-742-219-466

Query Match 8.3%; Score 70.5; DB 8; Length 899;
Best Local Similarity 23.9%; Pred. No. 7.8; Indels 53; Gaps 8;
Matches 39; Conservative 20; Mismatches 51; Indels 53; Gaps 8;
Qy 3 KKTIFYGAGWFTDRQNKAY-----KEAMEALKENPTIDL-----ENSYV 41
Db 324 KKYVWFLIGWADNWFKYDPSINCTVDEMTAEVGHITTEIVMLNPANTRISISNMTSQE 383
Qy 42 PLDNOYKQIRVDEHPE-----YLDKQWATATYNDLNGIKTNDIMLGVIYPDEED 92
Db 384 FVEKLTGK--RLKRHPETGGFQEAFLAYDAIWAALALN-----KTS----- 423
Qy 93 VGLGMELGYALSQGYVLLVLPDEDYDKPILMSW-GVSDNVI 134
Db 424 -GGGSGVRLDFNYYNQTTDQIY-RAMNSSSFEVSGHV 464

RESULT 14

US-60-732-162-406
; Sequence 406, Application US/60732162
; GENERAL INFORMATION:


```
; APPLICANT: Belouchi, Abdelmajid
; APPLICANT: Raelson, John V
; APPLICANT: Bradley, Walter E
; APPLICANT: Paquin, Bruno
; APPLICANT: Fournier, Helene
; APPLICANT: Nguyen-Huu, Quynh
; APPLICANT: Croteau, Pascal
; APPLICANT: Allard, Rene
; APPLICANT: Debrus, Sophie
; APPLICANT: Eerdewegh, Paul V
; APPLICANT: Little, Randall D
; APPLICANT: Keith, Tim
; APPLICANT: Segal, Jonathan
; TITLE OF INVENTION: Genemap of the Human Genes Associated With Asthma Disease
; FILE REFERENCE: 059908-5010-PR
; CURRENT APPLICATION NUMBER: US/60/732,162
; CURRENT FILING DATE: 2005-11-02
; NUMBER OF SEQ ID NOS: 4417
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 406
; LENGTH: 961
; TYPE: PR
; ORGANISM: Homosapiens
US-60-732-162-406

Query Match      8.3%; Score 70.5; DB 8; Length 961;
Best Local Similarity 23.9%; Pred. No. 8.5;
Matches 39; Conservative 20; Mismatches 51; Indels 53; Gaps 8;

Qy      3 KTIYFGAGWFTDRONKAY-----KEAMEALKENPTIDL-----ENSYV 41
Db      386 KKYVFLIGWYADNWFKIYDPSINCTVDTEAVEGHITTEIVMLNPANTRISNMTSQE 445
Qy      42 PLDNQYKGIKRVDEHPE-----YLDKQVWATATYNNDLNGIKTNDIMLGVIYPDEED 92
Db      446 FVEKLTGK--RLKRHPETGGFQEPAPLAYDAIWAALALN-----KTS----- 485
Qy      93 VGLGMELGYALSQKQYVLLVLPDSDYKPKINLMSW-GVSDNVI 134
Db      486 -GGGGRSGVRLEDFNYYNQITTDQIY-RAMNSSSFEFGVSGHVV 526

Search completed: December 23, 2005, 22:17:35
Job time : 15 secs

RESULT 15
US-60-742-219-460
; Sequence 460, Application US/60/42219
; GENERAL INFORMATION:
; APPLICANT: Belouchi, Abdelmajid
; APPLICANT: Raelson, John Verner
; APPLICANT: Bradley, Walter Edward
; APPLICANT: Paquin, Bruno
; APPLICANT: Nguyen-Huu, Quynh
; APPLICANT: Croteau, Pascal
; APPLICANT: Allard, Rene
; APPLICANT: Little, Randall David
; APPLICANT: Keith, Tim
; APPLICANT: Cousineau, Johanne
; APPLICANT: Eerdewegh, Paul Van
; APPLICANT: Segal, Jonathan
; TITLE OF INVENTION: Gene Map of the Human Genes Associated with Psoriasis
; FILE REFERENCE: 306522-2000
; CURRENT APPLICATION NUMBER: US/60/742,219
; CURRENT FILING DATE: 2005-12-05
; NUMBER OF SEQ ID NOS: 7303
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 460
; LENGTH: 961
; TYPE: PR
; ORGANISM: Homo sapiens
US-60-742-219-460

Query Match      8.3%; Score 70.5; DB 8; Length 961;
Best Local Similarity 23.9%; Pred. No. 8.5;
Matches 39; Conservative 20; Mismatches 51; Indels 53; Gaps 8;
```

```
Qy      3 KTIYFGAGWFTDRONKAY-----KEAMEALKENPTIDL-----ENSYV 41
Db      386 KKYVFLIGWYADNWFKIYDPSINCTVDTEAVEGHITTEIVMLNPANTRISNMTSQE 445
Qy      42 PLDNQYKGIKRVDEHPE-----YLDKQVWATATYNNDLNGIKTNDIMLGVIYPDEED 92
Db      446 FVEKLTGK--RLKRHPETGGFQEPAPLAYDAIWAALALN-----KTS----- 485
Qy      93 VGLGMELGYALSQKQYVLLVLPDSDYKPKINLMSW-GVSDNVI 134
Db      486 -GGGGRSGVRLEDFNYYNQITTDQIY-RAMNSSSFEFGVSGHVV 526
```

Search completed: December 23, 2005, 22:17:35
Job time : 15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2005, 21:50:25 ; Search time 38 Seconds
(without alignments)
397.527 Million cell updates/sec

Title: US-10-049-750-14
Perfect score: 848
Sequence: 1 MPKTIYFGAGWFTDRQNK.....QLKDFNFKPRDFYEGAVY 157

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	637	75.1	158	JC7522	nucleoside deoxyri
2	136.5	16.1	159	E86686	hypothetical prote
3	91	10.7	719	A42808	IG light chain-bin
4	90.5	10.7	660	G90330	amine oxidase (cop
5	86.5	10.2	462	A86717	conserved hypothet
6	83	9.8	206	B28439	maturase-related h
7	83	9.8	369	F84113	hypothetical prote
8	82.5	9.7	376	G71925	cag island protein
9	82.5	9.7	766	A40258	RAS GTPase-activat
10	82	9.7	1088	S39261	VPI protein - porc
11	82	9.7	1838	A42091	transcription acti
12	81.5	9.6	376	A64587	cag pathogenicity
13	81.5	9.6	2172	T00936	probable ATP-depen
14	81	9.6	241	H95078	hypothetical prote
15	81	9.6	313	S47433	cathepsin L (EC 3.
16	80.5	9.5	2233	S63347	acetyl-CoA carboxy
17	80	9.4	303	T17774	hypothetical prote
18	79.5	9.4	446	H97091	chitinase family p
19	79.5	9.4	1452	A97323	DNA polymerase III
20	78.5	9.3	236	H85435	hypothetical prote
21	78.5	9.3	423	E69936	conserved hypothet
22	78	9.2	410	C84827	hypothetical prote
23	78	9.2	751	HYRN	neprilysin (EC 3.4
24	78	9.2	816	A71006	hypothetical prote
25	77.5	9.1	816	C83917	DNA topoisomerase
26	77.5	9.1	2216	S78398	hypothetical prote
27	77	9.1	1088	P1XBRB	inner layer protei
28	77	9.1	1088	S13558	VPI protein - bovi
29	77	9.1	1650	T18444	hypothetical prote

30	76.5	9.0	241	2	E64476	hypothetical prote
31	76.5	9.0	432	2	G71695	proline-tRNA ligas
32	76	9.0	435	2	H75443	aspartyl-tRNA synt
33	76	9.0	455	2	A45063	immunoglobulin-bin
34	76	9.0	519	2	S45723	P60 protein - oat
35	76	9.0	2269	2	T28677	rhothy protein -
36	75.5	8.9	177	1	KIVZSW	thymidine kinase (
37	75.5	8.9	181	1	KIVZSK	hypothetical prote
38	75.5	8.9	256	2	D89769	hypothetical prote
39	75.5	8.9	437	2	B98074	probable TonB-depe
40	75.5	8.9	690	2	AB0124	hypothetical prote
41	75.5	8.9	759	2	T00153	hypothetical prote
42	75.5	8.9	1177	2	I64233	hypothetical prote
43	75.5	8.9	1260	2	H89984	hypothetical prote
44	75.5	8.9	1504	2	T17426	FK506 polyketide s
45	75	8.8	602	2	E64464	hypothetical prote

ALIGNMENTS

RESULT 1

JC7522
nucleoside deoxyribosyltransferase (EC 2.4.2.6) II - Lactobacillus helveticus
C:Species: Lactobacillus helveticus
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 09-Jul-2004
C:Accession: JC7522; PC7103
R:Okuyama, K.; Noguchi, T.
Biosci. Biotechnol. Biochem. 64, 2243-2245, 2000
A:Title: Molecular cloning and expression of the nucleoside deoxyribosyltransferase-II 9
A:Reference number: JC7522; MUID:21012342; PMID:11129605
A:Accession: JC7522
A:Molecule type: DNA
A:Residues: 1-158 <OKU>
A:Cross-references: UNIPROT:Q9KWF0; UNIPARC:UPI000015C9E3; DDBJ:AB039914
A:Experimental source: strain ATCC 8018
A:Accession: PC7103
A:Molecule type: protein
A:Residues: 1-11;53-62 <OK2>
A:Cross-references: UNIPARC:UPI000017A510; UNIPARC:UPI000017A511
C:Comment: This enzyme catalyzes transfer of glycosyl residues from a donor deoxynucleoside
C:Genetics:
A:Gene: ndtB
C:Keywords: glycosyltransferase; pentosyltransferase

Query Match	75.1%	Score 637;	DB 2;	Length 158;
Best Local Similarity	73.5%;	Pred. No. 8.9e-49;		
Matches 114;	Conservative 21;	Mismatches 20;	Indels 0;	Gaps 0;
QY	3	KKTIYFGAGWFTDRONKAYKEAMEALKENPTIDLENSVPLDNQYKGRVDEHPYLDHK	62	
Db	4	KKTIYFGAGWFTDRONKAYKEAMEALKENPTIDLENSVPLDNQYKGRVDEHPYLDHK	63	
QY	63	WVATATYNDLNGIKTNDIMLGVIIPDEEDVGLGELGVALSQKVVLLVIPDEYDGKPI	122	
Db	64	EWASATYNDLNGIKTNDIMLGVIIPDEEDVGLGELGVALSQKVVLLVIPDEYDGKPI	123	
QY	123	NLMMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY	157	
Db	124	ILMSGVCDNASQISELKDFDFNFKRYNFDGAVY	158	

RESULT 2

E86686
hypothetical protein vejD [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C:Accession: E86686
R:Boilotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s.
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: E86686


```
RESULT 6
B28439
maturase-related hypothetical protein RF2 - yeast (Saccharomyces cerevisiae) mitochondri
A:Species: Saccharomyces cerevisiae
A:Variety: Saccharomyces uvarum
C:Accession: B28439
C:Date: 28-Aug-1989 #sequence_revision 28-Aug-1989 #text_change 09-Jul-2004
R:Seraphin, B.; Simon, M.; Paye, G.
J. Biol. Chem. 262, 10146-10153, 1987
A:Title: The mitochondrial reading frame RF3 is a functional gene in Saccharomyces uvaru
A:Reference number: A32597; MUID:87280035; PMID:2440860
A:Accession: B28439
A:Molecule type: DNA
A:Residues: 1-101:102-206 <SER>
A:Cross-references: UNIPROT:Q7M2B0; UNIPARC:UPI0000093837; UNIPARC:UPI000017B3A7
A:Note: the source is designated as Saccharomyces uvarum
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC2
C:Keywords: mitochondrion

Query Match          9.8%; Score 83; DB 2; Length 206;
Best Local Similarity 21.8%; Pred. No. 4.7;
Matches 36; Conservative 28; Mismatches 51; Indels 50; Gaps 8;

Qy 9 GAGWFTDRONKAYKEAMEALKENPTIDLEN-----SYPLDNQYKGIKRVDEHPEYLH 60
Db 49 GDGVFGIKKNTHTSPNI-TLAKTDITITLENIKYILNLTSKIYVONKYNKVQ-----LH 100

Qy 61 DKVWATATYNNDLNGIKTNDIML-GVYIPDEEDVGLGMELGVALSQGYVLLVIPDEYDG 119
Db 101 TK-WFSTKIYINPIAFPMMALEIIIRGLYHTD-----KGNFMKLKDELNS 144

Qy 120 K--PINLSWGVSDNVIKMSLKDFN-----FNKP 147
Db 145 NLTDINHLPMWYSELLKMDPIYDLNMPHRMNCQYANSTRYKPK 189

RESULT 7
F84113
hypothetical protein BH3710 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C:Accession: F84113
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: F84113
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-369 <STO>
A:Cross-references: UNIPROT:Q9K6L9; UNIPARC:UPI00000C429C; GB:AP001519; GB:BA0000004; NID
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3710

Query Match          9.8%; Score 83; DB 2; Length 369;
Best Local Similarity 25.5%; Pred. No. 9.8;
Matches 25; Conservative 16; Mismatches 41; Indels 16; Gaps 4;

Qy 4 KTIYFGAGFTDRONKAYKEAMEALKEN-PTIDLENSVYPL-----DNQYKGIK-- 51
Db 190 KNFYLGIIHGVSPDDENKYNVKAAMESLANVYKENSVPVLLADKRNATSTQERKAIEYM 249

Qy 52 ---VDEHP-EYLHDKVWATATYNNDLNGIKTNDIMLGV 85
Db 250 SSLDDEEPIVYRHSNIWETCAFLSLDGVVTTKLHVG 287

RESULT 8
G71925
cag island protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: Strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C:Accession: G71925
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: G71925
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <ARN>
A:Cross-references: UNIPROT:Q9ZLU1; UNIPARC:UPI00000D363E; GB:AE001482; GB:AE001439; NID
A:Experimental source: strain J99
C:Genetics:
A:Gene: cagM

Query Match          9.7%; Score 82.5; DB 2; Length 376;
Best Local Similarity 26.0%; Pred. No. 11;
Matches 38; Conservative 22; Mismatches 71; Indels 15; Gaps 5;

Qy 17 QNKAYKEAMEALKENPTID--LENSYVPLDNQYKGIKRVDEHPEYLHDKVWATATYNNDL 73
Db 105 KSKAYGDLQALKDNPLYPYAVLVNQETFTQEDKERLSYYYPQVKTSSIFKKT 164

Qy 74 NGIKTND---IMLGVYIPDEEDVGLGMELGVALSQGYVLLVIPDEYKPI-NUMSWG 129
Db 165 ATTCKDAQALLQMGVFSLDEQNKKASRLALSXQA-----IEEYSNNISNLSRKE 216

Qy 130 SDNVIKMSQLKDFNFKRPFDFYEGA 155
Db 217 LDNIYVQLQERNKFDKAKDIAOKA 242

RESULT 9
A40258
RAS GTPase-activating protein earl - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 06-Dec-1991 #sequence_revision 06-Dec-1991 #text_change 09-Jul-2004
C:Accession: A40258; A56606; T40588
R:Imai, Y.; Miyake, S.; Hughes, D.A.; Yamamoto, M.
Mol. Cell. Biol. 11, 3088-3094, 1991
A:Title: Identification of a GTPase-activating protein homolog in Schizosaccharomyces po
A:Reference number: A40258; MUID:91246176; PMID:2038319
A:Accession: A40258
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-766 <IMA>
A:Cross-references: UNIPROT:P33277; UNIPARC:UPI000012B0D2; GB:D10457; NID:g218534; PIDN
R:Wang, Y.; Boguski, M.; Riggs, M.; Rodgers, L.; Wigler, M.
Cell Regul. 2, 453-465, 1991
A:Title: earl, a gene from Schizosaccharomyces pombe encoding a protein that regulates
A:Reference number: A56606; MUID:91355280; PMID:1883874
A:Accession: A56606
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-766 <WAN>
A:Cross-references: UNIPARC:UPI000012B0D2; GB:S37449; NID:g234782; PIDN:AAB19697.1; PID
A:Note: sequence extracted from NCBI backbone (NCBI:106678, NCBI:106679)
R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 1999
A:Reference number: Z21938
A:Accession: T40588
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-766 <SEE>
A:Cross-references: UNIPARC:UPI000012B0D2; EMBL:AL035216; PIDN:CAA22816.1; GSPDB:GN0006
A:Experimental source: strain 972h-; cosmid c646
C:Genetics:
A:Gene: SPBC646.12C
A:Map position: 2
```

F;154-398/Domain: ras-specific GAP catalytic domain homology <GAP>

Query Match 9.7%; Score 82.5; DB 2; Length 766;
Best Local Similarity 25.4%; Pred. No. 27;
Matches 46; Conservative 17; Mismatches 43; Indels 75; Gaps 10;

QY 24 AMEALKENPT---IDLNSYVP-----LDNQYKQIRVD-----EHPE 57

DB 562 AIELDELSTRLVVDKENRYEPTSEVEKEFIDLADYERIRABRDALQDVHRAICDNE 621

QY 58 YLHDKVMATATY--NNDLNGIKTNDIMLGVIYPDEEDVGLGMEGLYALSQGYVLLVIPDE 116

DB 622 YLQTLQIYGYLNNARSQIK-----PSHSD-----SKGFSRGVGVGI----- 660

QY 117 DYKQPINLMSWGVSDNVKMS--QLKD-----FNFNKPREFVEGAV 156

DB 661 ---KPKNIK-----SSNTVKLSQQLKESVLLNCTIPEFNVNTYFTTFSSPSTDNFVI 713

QY 157 Y 157

DB 714 Y 714

RESULT 10
S39261

VPI protein - porcine rotavirus

C;Species: porcine rotavirus

C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004

C;Accession: S39261

R;Almanza, L.; Arias, C.F.; Lopez, S.

submitted to the EMBL Data Library, November 1993

A;Description: Amino acid sequence of the Porcine rotavirus YM VP1 protein.

A;Reference number: S39261

A;Accession: S39261

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1088 <ALM>

A;Cross-references: UNIPROT:Q85036; UNIPARC:UPI0000061936; EMBL:X76486; NID:g434312; PII

C;Superfamily: rotavirus inner layer protein VPI

Query Match 9.7%; Score 82; DB 2; Length 1088;
Best Local Similarity 20.6%; Pred. No. 47;
Matches 35; Conservative 30; Mismatches 43; Indels 62; Gaps 8;

QY 16 RQNKAYKEAMEALKENP--TIDLENSYVPLDNOYKQIRVDPEHPEYLHDKVMATATYNN-- 71

DB 184 RHNEKYRYEYEVKMDKPYLVLTWANSATIE-----LMSVFSHEDYLLAKELLILSYNSRS 238

QY 72 -----DLNGIKTNDIMLGVIYPDEEDVGLGMEGLYALSQGYVLLVIP 114

DB 239 TLAKLVSSPMSILVALIDIN-----GTFITNE-----DELEFS---DKYVKAIVP 281

QY 115 DEDYKG-----PINLMSWGVSDNVK-----MSQLKDFNFN 145

DB 282 DQTFNELQEMIDNMKKAGLVDIRPMIQEWLVDCSLEKFTLMSKIYSWSFH 331

RESULT 11
A42091

transcription activator SNF2/SWI2 homolog brm - fruit fly (Drosophila melanogaster)

C;Species: Drosophila melanogaster

C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C;Accession: A42091

R;Tamkun, J.W.; Deuring, R.; Scott, M.P.; Kissinger, M.; Pattatucci, A.M.; Kaufman, T.C.

Cell 68, 561-572, 1992

A;Title: brhma: a regulator of Drosophila homeotic genes structurally related to the ye

A;Reference number: A42091; MUID:92154670; PMID:1346755

A;Accession: A42091

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1638 <TAM>

A;Cross-references: UNIPROT:P25439; UNIPARC:UPI0000016B03; GB:M85049; NID:g157011; PIDN:

A;Experimental source: iso-1

A;Note: sequence extracted from NCBI backbone (NCBIN:82354, NCBIIP:82360)
C;Genetics:

A;Gene: FlyBase:brm

A;Cross-references: FlyBase:FBgn0000212

C;Superfamily: human SNF2alpha protein; bromodomain homology

C;Keywords: transcription regulation

F;1451-1506/Domain: bromodomain homology <BRO>

Query Match 9.7%; Score 82; DB 2; Length 1638;
Best Local Similarity 20.2%; Pred. No. 79;
Matches 41; Conservative 30; Mismatches 46; Indels 86; Gaps 11;

QY 12 WFTDRQ-----NKAYKEAMEALKENPTI-----DLENSYVPLDNOYK 48

DB 688 WIDDESDSGSNDHHPKVE---EQPTATEDATDKAQTGNDKDLITKAKVEDDEY - 743

QY 49 GIRVDHPEY-----LHDKV-----WATATYNNDLNGIKTNDIM 82

DB 744 --RTEQTYYSIAHTIHKVVEQASIMVNGTLKEYQIKGLEWLVSYNNLNGILADEMG 801

QY 83 LG-----VYIPDEEDVGLGMEGLYALSQGYVLLV---IPD-----EDYGRPINL 124

DB 802 LGKTIQTISLVTYLMDRKV-----MGPLYLIIVPLSTLPNWLFEEKWAPAVGV 850

QY 125 MSWGVSDNVKIM--SQLKDFNFN 145

DB 851 VSYKQSGQGRRLQLQNMWRAKFN 873

RESULT 12

A64587

cag pathogenicity island protein cag16 - Helicobacter pylori (strain 26695)

C;Species: Helicobacter pylori

C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004

C;Accession: A64587

R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.

Nature 388, 539-547, 1997

A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.N.

A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A;Reference number: A64520; MUID:97394467; PMID:9252185

A;Accession: A64587

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-376 <TOM>

A;Cross-references: UNIPROT:O25270; UNIPARC:UPI00000D3000; GB:AE000568; GB:AE000511; NID

Query Match 9.6%; Score 81.5; DB 2; Length 376;
Best Local Similarity 25.3%; Pred. No. 14;
Matches 37; Conservative 23; Mismatches 71; Indels 15; Gaps 5;

QY 17 QNKAYKEAMEALKENPTID--LENSYVPLDNOYKQIRVD-EHPEYLHDKVMATATYNNDL 73

DB 105 KSKAYGDEALQKDNPLRYKLLPNPAYVLNQETFTTKEDRERLSYYPQVKTSSIFKTT 164

QY 74 NGIKTND---IMLGVIYPDEEDVGLGMEGLYALSQGYVLLVIPDEYGRKPI-NLMSWGV 129

DB 165 ATTCKAQALLQMGVFSLDSEQNKASRLSALYKQA-----IEEYSNNVSNLLSRKE 216

QY 130 SDNVKMSQLKDFNFNKPFDYEGA 155

DB 217 LDNIDYVYLQERNKFDKSKAKDIAQKA 242

RESULT 13

T00936

probable ATP-dependent RNA helicase At2g42270 [imported] - Arabidopsis thaliana

N;Alternate names: ATP-dependent RNA helicase homolog T24P15.18

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004

C;Accession: T00936; H84851

R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul,

```
submitted to the EMBL Data Library, December 1997
A:Description: Arabidopsis thaliana chromosome II BAC T24P15 genomic sequence.
A:Reference number: Z14212
A:Accession: T00936
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2172 <ROU>
A:Cross-references: UNIPROT:O48534; UNIPARC:UPI00000AB148; EMBL:AC002561; NID:g2673901;
A:Experimental source: cultivar Columbia
A:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-769, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: H84851
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1621, 'W', 1623-2172 <STO>
A:Cross-references: UNIPARC:UPI000017AP4A; GB:AE002093; NID:g2673917; PIDN:AA888651.1; G
C:Genetics:
A:Gene: T24P15.18; At2g42270
A:Map position: 2

Query Match          9.6%; Score 81.5; DB 2; Length 2172;
Best Local Similarity 23.2%; Pred. No. 1.3e+02;
Matches 39; Conservative 27; Mismatches 61; Indels 41; Gaps 8;

Qy 1 MPKTIYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVP-----LDNQYKGRVDEH 55
Db 435 MDLESAPNQGGFTRENNKC-----ELPDRFRIRGKEFDEHVFWVWSKKFDSNEKLKVISDL 492
Qy 56 PEY-----LHDKVWATATYND-----LNGIKTNDIMLGVTIPDEEDVGL 95
Db 493 PEWAQPAFRGQQLNRVQSVGYTALFKADNILLCAPTGAGTINVAVTIL-----HQLGL 548
Qy 96 GNEGLYALSQGYVLLVTPDEYDKPINLMSGVSDNVKMSQ-LKDF 142
Db 549 NNPFGTTFHNGYKIV-----YVAPMKAL---VAEVVDSLSQRLKDF 587

RESULT 14
H95078
hypothetical protein SP0680 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C:Accession: H95078
R:Retelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: H95078
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-241 <KUR>
A:Cross-references: UNIPROT:Q97RV6; UNIPARC:UPI000005153B; GB:AE005672; PIDN:AAK74825.1;
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0680
C:Superfamily: conserved hypothetical protein H11243

Query Match          9.6%; Score 81; DB 2; Length 241;
Best Local Similarity 24.9%; Pred. No. 8.5;
Matches 47; Conservative 19; Mismatches 59; Indels 64; Gaps 10;

Qy 13 FTDRQNKAYKEAMEAL-----KENPTI-DLENSYVPLDNQYKGRVDE----- 54
Db 49 FODRIIQGYEHTYLMHLHPAGAVTANKKELFTVMDLLPSNIQSKLYAVAGRLDRDTGL 108
Qy 55 -----HPEYLHDKVWATATYNNDLNGIKTND-----IMLGVIYPDE----- 90
```

```
Db 109 LLLTDNGPLGFLQLHPQHVDK-----TYQEVNGLLTPDHIQTFOKGI VFLDDTVCKPA 163
Qy 91 --EDVGLGMELGYA---LSQGYVLLVTPDEYDKPINLMSGVSDNVKMSQLKDFNFN 145
Db 164 KLEILSASPSLSQASITISEGKFQI-----KKM-FLSVGVKVTSLKRIQFGDFTLN 214
Qy 146 KPRDFYFEG 154
Db 215 P---DLAEG 220

RESULT 15
S47433
cathepsin L (EC 3.4.22.15) - Norway lobster
C:Species: Nephrops norvegicus (Norway lobster)
C:Date: 23-Nov-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C:Accession: S47433
R:le Boulay, C.; van Wormhoudt, A.; Sellos, D.
submitted to the EMBL Data library, August 1994
A:Description: Molecular cloning and sequencing of the two cDNAs that encode cathepsin L
A:Reference number: S47432
A:Accession: S47433
A:Molecule type: mRNA
A:Residues: 1-313 <LEB>
A:Cross-references: UNIPROT:Q27708; UNIPARC:UPI0000084019; EMBL:X80990; NID:g530735; PID
C:Superfamily: papain
C:Keywords: cysteine proteinase; hydrolase
F:122,260,280/Active site: Cys, His, Asn #status predicted

Query Match          9.6%; Score 81; DB 2; Length 313;
Best Local Similarity 26.3%; Pred. No. 12;
Matches 44; Conservative 15; Mismatches 48; Indels 60; Gaps 11;

Qy 9 GAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNQYKGRVDP-----E 54
Db 163 GGGWMT-----SAFDYIKDNGIDTSSY-PYEAODRSCRFPDANSIGATCTGFEVQ 213
Qy 55 H-PYLLHDKVWATATYNNDLNGIKT-----NDIMLGVIY-----PDEEDVGLGMEL 99
Db 214 HTEERALHEAV-----SDIGPISVAIDASHFSPQFYSSGVYKCKSPTNLDHGV-LAV 265
Qy 100 GYALSQGYVLLVTPDEYDKPINLMSGV---SDNVKMSQLKDFN 143
Db 266 GYGTES-----TEDYWLVKN--SWGSGMGDAGYIKMSRNRDNN 301

Search completed: December 23, 2005, 22:04:55
Job time : 42 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2005, 21:48:35 ; Search time 230 Seconds
(without alignments)
481.600 Million cell updates/sec

Title: US-10-049-750-14
Perfect score: 848
Sequence: 1 MPKTYIFGAGWFTDRQNK.....QLKDFNFNKRDFEYEGAVY 157

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	843	99.4	156	1 NTD_LACIE	Q9r5v5 lactobacill
2	728	85.8	158	2 Q8RLY4 LACHE	Q8rlY4 lactobacill
3	719	84.8	159	2 Q5FMH8 LACAC	Q5fmH8 lactobacill
4	653	77.0	149	1 NTD_LACJO	Q74lq9 lactobacill
5	634	74.8	158	2 Q9KWF0 LACHE	Q9kwF0 lactobacill
6	209	24.6	168	1 NTD_LACFE	Q6vni5 lactobacill
7	171.5	20.2	167	2 Q8RLY5 LACHE	Q8rlY5 lactobacill
8	154.5	18.2	167	2 Q5FIN0 LACAC	Q5fin0 lactobacill
9	136.5	16.1	159	2 Q9CI73 LACLA	Q9ci73 lactococcus
10	124.5	14.7	157	2 Q88SL4 LACPL	Q88sl4 lactobacill
11	106	12.5	155	2 Q4QAY0 LEIMA	Q4qay0 leishmania
12	96	11.3	146	2 Q88X23 LACPL	Q88x23 lactobacill
13	96	11.3	1149	2 Q5EMR0 AMEPV	Q5emr0 amacta moo
14	95	11.2	1029	2 Q5ZMQ1 CHICK	Q5zmQ1 gallus gall
15	93.5	11.0	264	2 Q4Z0Q6 PLABE	Q4z0Q6 plasmodium
16	93.5	11.0	325	2 Q7P6V5 FUSNV	Q7p6V5 fusobacteri
17	92.5	10.9	800	2 Q4L5E4 STAHJ	Q4l5e4 staphylococ
18	92	10.8	553	1 ADSC_METAC	Q8trh5 methanosarc
19	91	10.7	719	2 Q5I9I2 PEPMA	Q5i9I2 peptostrept
20	90.5	10.7	660	2 Q97XMI SULSO	Q97xmI sulfolobus
21	90	10.6	454	2 Q4JUB7 CORJK	Q4jub7 corynebacte
22	89	10.5	218	2 Q8RHH4 FUSNV	Q8rhh4 fusobacteri
23	89	10.5	462	2 Q7WST1_PSEPA	Q7wst1 pseudomonas
24	89	10.5	844	2 Q7R980_PLAYO	Q7r989 plasmodium
25	89	10.5	1139	2 Q7RRP6_PLAYO	Q7rrp6 plasmodium
26	88	10.4	504	1 AMPA_CANBF	Q7vqt0 candidatus
27	88	10.4	544	2 Q7P5F2 FUSNV	Q7p5f2 fusobacteri
28	87.5	10.3	519	2 Q8RT73_PLAYO	Q8rt73 plasmodium
29	87	10.3	678	2 Q8TMR7_METAC	Q8tmr7 methanosarc
30	86.5	10.2	462	2 Q9CHJ4_LACIA	Q9chj4 lactococcus
31	85.5	10.1	438	2 Q9H804_HUMAN	Q9h804 homo sapien

RESULT 1

ID	NTD_LACIE	STANDARD	PRT	156 AA
AC	Q9R5V5			
DT	25-OCT-2004 (Rel. 45, Last sequence update)			
DT	25-OCT-2004 (Rel. 45, Last sequence update)			
DT	01-FEB-2005 (Rel. 46, Last annotation update)			
DE	Nucleoside deoxyribosyltransferase (EC 2.4.2.6) (N-deoxyribosyltransferase).			
GN	Name=ntd;			
OS	Lactobacillus leichmannii.			
OC	Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;			
OC	Lactobacillus.			
OX	NCBI_TaxID=28039;			
RN	[1]			
RP	NUCLEOTIDE SEQUENCE, PROTEIN SEQUENCE OF 1-25, FUNCTION, ACTIVE SITE, AND MUTAGENESIS OF GLU-97.			
RX	MEDLINE=95318137; PubMed=7797550; DOI=10.1074/jbc.270.26.15551;			
RA	Porter D.J.T., Merrill B.M., Short S.A.;			
RT	"Identification of the active site nucleophile in nucleoside 2-deoxyribosyltransferase as glutamic acid 98.";			
RL	J. Biol. Chem. 270:15551-15556(1995).			
RN	[2]			
RP	X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).			
RX	MEDLINE=96419140; PubMed=8805514;			
RA	Armstrong S.R., Cook W.J., Short S.A., Balick S.E.;			
RT	"Crystal structures of nucleoside 2-deoxyribosyltransferase in native and ligand-bound forms reveal architecture of the active site.";			
RL	Structure 4:97-107(1996).			
CC	FUNCTION: Catalyzes the cleavage of the glycosidic bond of 2'-deoxyribonucleosides and the transfer of the deoxyribosyl moiety to an acceptor purine or pyrimidine base.			
CC	CATALYTIC ACTIVITY: 2-deoxy-D-ribose(1) + base(2) = 2-deoxy-D-ribose(1)-base(2) + base(1).			
CC	BIOPHYSICOCHEMICAL PROPERTIES:			
CC	pH dependence: Optimum pH is 6.0;			
CC	PATHWAY: Nucleotide metabolism.			
CC	SUBUNIT: Homohexamer.			
CC	SIMILARITY: Belongs to the nucleoside deoxyribosyltransferase family.			
CC	CATION: Was originally (Ref.1) thought to originate from E.coli.			
CC	This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.			
CC	PDB: 1F8X; X-ray; A/B=1-156.			
CC	PDB: 1F8Y; X-ray; A/B=1-156.			
DR	InterPro; IPR007710; N deoxyrib trans.			
DR	Pfam; PF05014; Nuc_deoxyrib_tr; 1.			
KW	3D-structure; Direct protein sequencing; Nucleotide metabolism;			


```

KW Transferase.
FT INIT_MET 0 0 Nucleophile.
FT MUTAGEN 97 97 E->A: Loss of transferase activity.
SQ SEQUENCE 156 AA; 17949 MW; 3A3ABC3F5B84743B CRC64;

Query Match 99.4%; Score 843; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. NO. 4.6e-64;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PKTIYFGAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNOYKGIRVDEHPEYLHD 61
DB 1 PKTIYFGAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNOYKGIRVDEHPEYLHD 60
QY 62 KWATATYNNDLNGIKTNDIMLGVIYPDEEDVGLGMELGYALSQGYKYLVIIPDEDYGRP 121
DB 61 KWATATYNNDLNGIKTNDIMLGVIYPDEEDVGLGMELGYALSQGYKYLVIIPDEDYGRP 120
QY 122 INLMSWGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157
DB 121 INLMSWGVSDNVIKMSQLKDFNFKPRDFYEGAVY 156

RESULT 2
Q8RLY4_LACHE PRELIMINARY; PRT; 158 AA.
AC Q8RLY4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 24, Last annotation update)
DE N-deoxyribosyltransferase.
GN Name=ntd;
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1587;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CNR232;
RX MEDLINE=21964041; PubMed=11836245; DOI=10.1074/jbc.M111995200;
RA Kaminaki P.A.;
RT "Functional cloning, heterologous expression, and purification of two
different N-deoxyribosyltransferases from Lactobacillus helveticus.";
RL J. Biol. Chem. 277:14400-14407(2002).
DR EMBL; AY064167; AAL73114.1; -; Genomic_DNA.
DR HSP; Q9RSV5; 1F8Y.
DR SMR; Q8RLY4; 4-158.
DR GO; GO:0016740; P:transferase activity; IEA.
DR InterPro; IPR007710; N_deoxyrib_trans.
DR Pfam; PF05014; Nuc_deoxyrib_tr; 1.
KW Transferase.
SQ SEQUENCE 158 AA; 18148 MW; 00FC9E4B210C47B8 CRC64;

Query Match 85.8%; Score 728; DB 2; Length 158;
Best Local Similarity 82.6%; Pred. No. 3.1e-54;
Matches 128; Conservative 20; Mismatches 7; Indels 0; Gaps 0;

QY 3 KTIYFGAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNOYKGIRVDEHPEYLHDK 62
DB 4 KTIYFGAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLENQGIRIDEHPEYLHNI 63
QY 63 VWATATYNNDLNGIKTNDIMLGVIYPDEEDVGLGMELGYALSQGYKYLVIIPDEDYKPI 122
DB 64 EWASATYHNDLIGIKTSVLMGVYLPDEEDVGLGMELGYALSQGYKYLVIIPDEDYKPI 123
QY 123 NLMSWGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157
DB 124 NLMSWGVCDNAIKSELKDFNFKPRYFNFDGAVY 158

RESULT 3
Q5FMM8_LACAC PRELIMINARY; PRT; 159 AA.
ID Q5FMM8_LACAC PRELIMINARY;

```

```

AC Q5FMM8;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE 2-deoxyribosyltransferase.
GN OrderedLocusNames=LBA0145;
DB Lactobacillus acidophilus.
OS Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1579;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCFM;
RX PubMed=15671160; DOI=10.1073/pnas.0409188102;
RA Altermann E., Russell W.M., Azcarate-Peril M.A., Barrangou R.,
RA Buck B.L., McAuliffe O., Souther N., Dobson A., Duong T., Callanan M.,
RA Lick S., Hanrick A., Cano R., Klaenhammer T.R.;
RT "Complete genome sequence of the probiotic lactic acid bacterium
Lactobacillus acidophilus NCFM.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:3906-3912(2005).
DR EMBL; CP000033; AAV42046.1; -; Genomic DNA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR007710; N_deoxyrib_trans.
DR Pfam; PF05014; Nuc_deoxyrib_tr; 1.
KW Complete proteome; Transferase.
SQ SEQUENCE 159 AA; 18252 MW; 5CC7FFDBA57938FE CRC64;

Query Match 84.8%; Score 719; DB 2; Length 159;
Best Local Similarity 83.1%; Pred. No. 1.8e-53;
Matches 128; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

QY 4 KTIYFGAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNOYKGIRVDEHPEYLHDKV 63
DB 6 KTIYFGAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLENQGIRIDEHPEYLHDIE 65
QY 64 WATATYNNDLNGIKTNDIMLGVIYPDEEDVGLGMELGYALSQGYKYLVIIPDEDYKPIN 123
DB 66 WASATYHNDLIGIKSSDIMLGVIYPDEEDVGLGMELGYALSQGYKYLVIIPDEDYKPIN 125
QY 124 LMSWGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157
DB 126 LMSWGVCDNAIKSELKDFNFKPRYFNFDGAVY 159

RESULT 4
NTD_LACJO STANDARD; PRT; 149 AA.
ID NTD_LACJO
AC Q74LQ9;
DT 25-OCT-2004 (Rel. 45, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Nucleoside deoxyribosyltransferase (EC 2.4.2.6) (N-
deoxyribosyltransferase).
GN Name=ntd; OrderedLocusNames=LJ0124;
OS Lactobacillus johnsonii.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=33959;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=NCC 533;
RX PubMed=14983040; DOI=10.1073/pnas.0307327101;
RA Pridmore R.D., Berger B., Desiere F., Vilanova D., Barretto C.,
RA Pittet A.-C., Zwahlen M.-C., Rouvet M., Altermann E., Barrangou R.,
RA Mollet B., Mercenier A., Klaenhammer T., Arigoni F., Schell M.A.;
RT "The genome sequence of the probiotic intestinal bacterium
Lactobacillus johnsonii NCC 533.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:2512-2517(2004).
CC -I- FUNCTION: Catalyzes the cleavage of the glycosidic bond of 2'-
deoxyribonucleosides and the transfer of the deoxyribosyl moiety
to an acceptor purine or pyrimidine base (By similarity).
CC -I- CATALYTIC ACTIVITY: 2-deoxy-D-riboseyl-base(1) + base(2) = 2-deoxy-
D-riboseyl-base(2) + base(1).

```

QY 63 VVATATYNDNLGKIKTNDIMLVYIPDEEDVGLGMEIGYALSOQKYVLLVIPDEDYCKPI 122
 DB 64 EWASATYHNDLVGKITSVDLLGVYLQFEHVHGLGMEIGYALSOQKFFWFHSHMKDYCKPI 123

QY 123 NLMWSGVSDNVIKVQSKQKDFNFKPRFDFYEGAVY 157
 DB 124 ILMWSGVCDNASQISELKDFDFNFKPRYFYDGAVY 158

RESULT 6
 NTD_LACFE STANDARD; PRT; 168 AA.
 AC QGVN15;
 DT 25-OCT-2004 (Rel. 45, Created)
 DT 25-OCT-2004 (Rel. 45, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)
 DE Nucleoside deoxyribosyltransferase (EC 2.4.2.6) (N-
 deoxyribosyltransferase).
 GN Name:ntd;
 OS Lactobacillus fermentum.
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
 OC Lactobacillus.
 NCBI_TaxID=1613;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
 RC STRAIN=CIP 102980T;
 RA Kaminski P.A.;
 RT Characterization of the Lactobacillus fermentum N-
 deoxyribosyltransferase.";
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Catalyzes the cleavage of the glycosidic bond of 2'-
 deoxyribonucleosides and the transfer of the deoxyribosyl moiety
 to an acceptor purine or pyrimidine base [By similarity].
 CC -!- CATALYTIC ACTIVITY: 2-deoxy-D-ribose(1)-base(1) + 2-deoxy-
 D-ribose(1)-base(2) + base(1).
 CC -!- PATHWAY: Nucleotide metabolism.
 CC -!- SIMILARITY: Belongs to the nucleoside deoxyribosyltransferase
 family.

 This Swiss-Prot entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use as long as its content is in no way modified and this statement is not
 removed.

 CC EMBL; AY064168; AAL73115.1; -; Genomic DNA.
 CC InterPro; IPR007710; N_deoxyrib_trans.
 CC Pfam; PF05014; Nuc_deoxyrib_tr; 1.
 CC Nucleotide metabolism; Transferase.
 FT ACT SITE 103 103 Nucleophile (By similarity).
 SQ SEQUENCE 168 AA; 18896 MW; 362162F43586C317 CRC64;

 Query Match 24.6%; Score 209; DB 1; Length 168;
 Best Local Similarity 36.4%; Pred. No. 6.8e-10;
 Matches 55; Conservative 23; Mismatches 59; Indels 14; Gaps 7

QY 6 IYFGAGFTDRQNKAYKEAMEALKENPTIDLENSVPLDNOYKGRVDEHPEYLDHKV-W 64
 DB 12 IYLATSPFNEQRARIPOALQLEANTVGVVHQ--PFDFQYKDARVSDPAGVFGSLEW 69

QY 65 ATATYNDNLGKIKTNDIMLVYIPDEEDVGLGMEIGYALSOQK-YVLLVIPDEDYCK-PI 122
 DB 70 QIATYNDNLNAVGTSDVCVALYDMDQIDEGICMEIGMEFVALHKPIVLLPFTKKDKSAYEA 129

QY 123 NLM-----SWGVSDNVIKMSQLKDFNFKP 147
 DB 130 NLMLAGVTTW-LEPN--DFSLPKDFNFNHP 157

RESULT 7
 Q8RLY5 LACHE PRELIMINARY; PRT; 167 AA.
 AC Q8RLY5;

```
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Purine trans deoxyribosylase (Nucleoside deoxyribosyltransferase-I)
DE (EC 2.4.2.6).
DE Name=ptd; Synonym=sandtA;
GN Lactobacillus helveticus.
OS Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1587;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CRZ32;
RX MEDLINE=21964041; PubMed=11836245; DOI=10.1074/jbc.M11995200;
RA Kaminski P.A.;
RT "Functional cloning, heterologous expression, and purification of two
RT different N-deoxyribosyltransferases from Lactobacillus helveticus.";
RL J. Biol. Chem. 277:14400-14407(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Okuyama K., Noguchi T.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY064166; AAL73113.1; -; Genomic DNA.
DR EMBL; AB076265; BAC00952.1; -; Genomic DNA.
DR PDB; 1S2D; X-ray; A/B/C=-.
DR PDB; 1S2G; X-ray; A/B/C=1-167.
DR PDB; 1S2I; X-ray; A/B/C=1-167.
DR PDB; 1S2L; X-ray; A/B/C=1-167.
DR PDB; 1S3F; X-ray; A/B/C=1-167.
DR GO; GO:0050144; F:nucleoside deoxyribosyltransferase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR InterPro; IPR007710; N deoxyrib trans.
DR Pfam; PF05014; Nuc deoxyrib tr; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 167 AA; 18713 MW; 5540581511CB4787 CRC64;

Query Match 20.2%; Score 171.5; DB 2; Length 167;
Best Local Similarity 31.4%; Pred. No. 1.1e-06;
Matches 49; Conservative 34; Mismatches 58; Indels 15; Gaps 7;

QY 1 MPKTIYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDQYKGRVDEHPEY-- 58
DB 5 VPTGKIYLGFFYSDAQREARAKAKELLAKNPIS1--AHVFFPFDGDTD-PDEKNPEIGG 61
QY 59 LHKVWATATYNNDLNGIKTNDIMLGVIYIPDEEDVGLGMELGYALSQGYVLLV---IP 114
DB 62 IRSVWRDATYQNDLTGTSNATCGVFLYDMQDLDGSAFSGFWRAMHKPVILVPFTEHP 121
QY 115 DEDYGKPINLM-SWGVS---DNVTKMSQLKDFNFK 146
DB 122 EKE--KKQNLMIAGQVTTIIDGNTEPEKLDYFNFE 155

RESULT 8
QSFNO_LACAC
ID QSFNO_LACAC PRELIMINARY; PRT; 167 AA.
AC QSFNO;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Putative deoxyribosyltransferase.
GN OrderedLocuNames=LBAL631;
OS Lactobacillus acidophilus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1579;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCFM;
RX PubMed=15671160; DOI=10.1073/pnas.0409188102;
RA Altermann E., Russell W.M., Azcarate-Peril M.A., Barrangou R.;
RA Buck B.L., McAuliffe O., Southern N., Dobson A., Duong T., Callanan M.,
RA Lick S., Hamrick A., Cano R., Klaenhammer T.R.;

RT "Complete genome sequence of the probiotic lactic acid bacterium
RT Lactobacillus acidophilus NCFM.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:3906-3912(2005).
DR EMBL; CP000033; AA043444.1; -; Genomic DNA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR007710; N deoxyrib trans.
DR Pfam; PF05014; Nuc deoxyrib tr; 1.
KW Complete proteome; Transferase.
SQ SEQUENCE 167 AA; 18933 MW; A80408F06C3E4D98 CRC64;

Query Match 18.2%; Score 154.5; DB 2; Length 167;
Best Local Similarity 32.5%; Pred. No. 3e-05;
Matches 49; Conservative 25; Mismatches 60; Indels 17; Gaps 7;

QY 6 IYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDQYKGRVDEHPEY--LH 60
DB 10 IYLGTFYNDQDQARVKKARALLEQNTV--VRVHEFFQNF---VDPEKDPPEADGLR 63
QY 61 DKVWATATYNNDLNGIKTNDIMLGVIYIPDEEDVGLGMELGYALSQGYVLLV--IPEDY 118
DB 64 SMTWRLATYNNDLSGIINATCGVFLYDMQDLDGSAFSGFWRAMHKPVILVPFTNDPNK 123
QY 119 GKPINLM-SWGVS---DNVTKMSQLKDFNFK 145
DB 124 EKQNLMIAGQVTTIIDGNTEPEKLDYFNFE 154

RESULT 9
Q9C173_LACLA
ID Q9C173_LACLA PRELIMINARY; PRT; 159 AA.
AC Q9C173;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein yefJb.
GN Names=yefJb; OrderedLocuNames=LL0493;
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471; DOI=10.1101/gr.GR-1697R;
RA Bolotin A., Wincker P., Mauder S., Jaillon O., Malarme K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
DR EMBL; AS006284; AAK04591.1; -; Genomic DNA.
DR PIR; E86686; E86686.
DR InterPro; IPR007710; N deoxyrib trans.
DR Pfam; PF05014; Nuc deoxyrib tr; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 159 AA; 18295 MW; 98AE589D9C099C3 CRC64;

Query Match 16.1%; Score 136.5; DB 2; Length 159;
Best Local Similarity 26.7%; Pred. No. 0.00099;
Matches 43; Conservative 34; Mismatches 57; Indels 27; Gaps 7;

QY 6 IYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDQYKGRVDEHPEYLDKVV-- 63
DB 11 VYLADPFSESQIKVLELLENALSKNTV--ANFFSPMRCQ-----HPESLPQVEA 60
QY 64 ----WATATYNNDLNGIKTNDIMLGVIYIPDEEDV--GLGMELGYALSQGYVLLVDPED 117
DB 61 FTPEWAKATMENDVNEVKNADIIVAFDHDQDTSGTAWELGYATALEKPTYLIRFEDT 120
QY 118 YGKPINLMSWGVSDN----VTKMSQLKDFNFKPRDFYEG 154
DB 121 I--PANIM--LTERNRAFTQIEQVEEYDFLESKLIPYSG 156

RESULT 10
```

```

Q88SL4_LACPL
ID Q88SL4_LACPL PRELIMINARY; PRT; 157 AA.
AC Q88SL4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein lp_3397.
GN OrderedLocustNames=lp_3397;
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCIMB 8826 / WCFS1;
RX MEDLINE=22480296; PubMed=12566566; DOI=10.1073/pnas.0337704100;
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hofer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
RA De Vos W.M., Siezen R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCFS1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
DR EMBL; AL935262; CAD65499.1; -; Genomic_DNA.
DR InterPro; IPR007710; N_deoxyrib_trans.
DR Pfam; PF05014; Nuc_deoxyrib_tr; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 157 AA; 17604 MW; 15E1PB38FE1F6746 CRC64;

Query Match 14.7%; Score 124.5; DB 2; Length 157;
Best Local Similarity 24.6%; Pred. No. 0.01; Mismatches 50; Indels 15; Gaps 4;
Matches 31; Conservative 30;

Qy 6 IYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYKGIKIRVDEHPEYLHDKVWA 65
Db 4 VYLAAPFDEAQRKIQKVSALLANPTINDGIFPEHQF----EERP-FGSRAMQ 56

Qy 66 TATYNNDLNGIKTNDIMLGVYIP-----IPDEEDVGLGMELGYALSQGYKVVLLVDPDEY 119
Dy 57 QYVYASDRQVRADVVVAIILDFMTSATNSPDGTMFEIGAARVAEKTPTVIIV--QFDAN 114

Qy 120 KPINLM 125
Dy 115 KELNLM 120

RESULT 11
Q04QAY0_LEIMA
ID Q04QAY0_LEIMA PRELIMINARY; PRT; 155 AA.
AC Q04QAY0;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=LmjF23.1580;
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Friedlin;
RA Peacock C.S., Murphy L., Ivens A.C., Berriman M., Blackwell J.,
RA Smith D., Collins M., Foster N., Harris D., Oliver K., O'Neill S.,
RA Saunders D., Seeger K., Warren T., Apostolou Z., Bauser C., Beck A.,
RA Bianchetti G., Borzym K., Bothe G., Bruch C., Ciaroni L.,
RA Dueterhoeft A., Fuchs M., Gabel C., Goffeau A., Hilbert H.,
RA Klages S., Kube M., Marra M., Masuy D., Mueller-Auer S., Pohl T.,
RA Purnelle B., Reinhardt R., Rieger M., Robben J., Schaefer M.,
RA Tosato V., Volckaert G., Wambutt R., Wedler H., Zimmermann M.,
RA Rajandream M., and Barrell B.G.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR005262; CAJ05145.1; -; Genomic_DNA.
KW Hypothetical protein.

SQ SEQUENCE 155 AA; 16887 MW; A9C053404F8D13F0 CRC64;

Query Match 12.5%; Score 106; DB 2; Length 155;
Best Local Similarity 27.6%; Pred. No. 0.39;
Matches 40; Conservative 16; Mismatches 47; Indels 42; Gaps 6;

Qy 4 KTIYF-GAGWFTDRQNKAYKEAMEA-LKENPTIDLENSYVPLDNOYKGIKIRVDEHPEYLHDK 61
Dy 5 KTIYIAGPAVHPDNGEAYNNHVRALLKEKGVPL-----IPVDNIATG----- 48

Qy 62 KMWATATYNNDLNGIKTNDIMLGVYIP-----DEEDVGLGMELGYALSQGYKVVLLVDP--- 115
Dy 49 ---ALSRKNKIDMIRACDAVIADLSFSPKSPDCGTAFELGYAALGKLLLTFTTTRP 105

Qy 116 -----EDYKPINLM 125
Dy 106 MVEYKGYGEMAEGLSVENFGPLFNLM 130

RESULT 12
Q88X23_LACPL
ID Q88X23_LACPL PRELIMINARY; PRT; 146 AA.
AC Q88X23;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein lp_1427.
GN OrderedLocustNames=lp_1427;
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCIMB 8826 / WCFS1;
RX MEDLINE=22480296; PubMed=12566566; DOI=10.1073/pnas.0337704100;
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hofer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
RA De Vos W.M., Siezen R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCFS1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
DR EMBL; AL935256; CAD63892.1; -; Genomic_DNA.
DR InterPro; IPR007710; N_deoxyrib_trans.
DR Pfam; PF05014; Nuc_deoxyrib_tr; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 146 AA; 16608 MW; E33D94B52F7BE2A9 CRC64;

Query Match 11.3%; Score 96; DB 2; Length 146;
Best Local Similarity 25.3%; Pred. No. 2.6;
Matches 39; Conservative 28; Mismatches 63; Indels 24; Gaps 8;

Qy 6 IYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYKGIKIRVDEHPEYLHDKVWA 65
Dy 4 IYIASPFSPPEQVTRVQLAEALQNPTV--TDYISPLHQ-----DAQEQF-TKTWA 54

Qy 66 TATYNNDLNGIKTNDIMLGVYIPDEEDVGLGMELGYALSQGYKVVLLVDPDEYKGPIN 123
Dy 55 TEIFHRDWAQIAAAQVIVTVLDPEAKNLDSTAYELGVATMRD---LPIIALQEKDEAVN 111

Qy 124 LM-----SWGVSNDNVIKMSQLKDFNFK-PRPFD 151
Dy 112 LMITESLHWYTK-----QVSDPETYDFQQLPKGDF 141

RESULT 13
Q9EMR0_AMEPV
ID Q9EMR0_AMEPV PRELIMINARY; PRT; 1149 AA.
AC Q9EMR0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

```

```

DE Name=AMV139;
GN Anasacta moorei entomopoxvirus (AmEPV).
OS Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Betaentomopoxvirus.
OX NCBI_TaxID=28321;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20396580; PubMed=10936094; DOI=10.1006/viro.2000.0449;
RA Bawden A.L., Glassberg K.J., Diggins J., Shaw R., Farmerie W.,
RA Moyer R.W.;
RT "Complete genomic sequence of the Anasacta moorei entomopoxvirus:
RT analysis and comparison with other poxviruses.";
RL Virology 274:120-139(2000).
DR EMBL; AF250284; AAGO2845.1; -; Genomic DNA.
SQ SEQUENCE 1149 AA; 135479 MW; 7E4EE54A13F87691 CRC64;

Query Match 11.38; Score 96; DB 2; Length 1149;
Best Local Similarity 23.74; Pred. No. 27;
Matches 42; Conservative 29; Mismatches 62; Indels 44; Gaps 10;

QY 13 FTD--RONKAYKEAMEALKENPTIDLEN-----SYVPLD-----NOYKGIKRVDEH 55
DB 12 FTFEIRNLFPRLNLSKKNENIINNITYKIEELFKYIYTHPLDLLTIRDISNADKDEY 71

QY 56 PEYLHDKVWATATYNN-DL--NGIKTNDIMLGV-----YIPDEEDVGLGMELGYALSQ-- 105
DB 72 VKQPVNNLYLRAYNEMDFKNIRYDDKVSIIINEINYPPEHTSBELKYRLSHYESES 131

QY 106 ---GKYVLLV-IPDEDYG-----KPINLMSWGSVDNVKMSQLKDNFNKPRDFY 152
DB 132 IRGRGVVTFSGVPDNGYGYLLSQSDPSKYSKIWAIVDNYLMID-----NEDKDFDY 181

RESULT 14
Q5ZMQ1.CHICK PRELIMINARY; PRT; 1029 AA.
AC Q5ZMQ1;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Hypothetical protein.
GN ORFNames=KCUMB04_1h14;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CB; TISSUE=Bursa;
RA Caldwell R.B., Kierzek A.M., Arakawa H., Bezzubov Y., Zaim J.,
RA Fiedler P., Kuter S., Blagodatski A., Kostovska D., Koter M.,
RA Plachy J., Carninci P., Hayashizaki Y., Buerstedde J.M.;
RT "Full-length cDNAs from chicken bursa lymphocytes to facilitate
RT gene function analysis.";
RL Genome Biol. 6:R6-R6(2005).
DR EMBL; AJ719333; CAG30992.1; -; mRNA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR012961; DSHCT.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF08148; DSHCT; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
KW Hypothetical protein.
SQ SEQUENCE 1029 AA; 116048 MW; EC8909BE178AF562 CRC64;

```

```

Query Match 11.2%; Score 95; DB 2; Length 1029;
Best Local Similarity 22.4%; Pred. No. 29;
Matches 45; Conservative 32; Mismatches 56; Indels 68; Gaps 9;

QY 1 MPKKTIVF-----GAGWFTDRQNKAYKEAMEAL-----KENPTID-- 35
DB 478 MPARTVLTFTSASKDFKDFRWISSGEYIQMSGRAGRRGMDDRGIVILMVDERKMSPTIGKQ 537

QY 36 -LENSVYPLDNOY-----KGIKRVDE-HPEYLHDKVWATATYNNDLNGI-----KTND 80
DB 538 LLKGSADPLNSAFHLTYNNVNLRLRVEEINPEYMLEKSFYQFOHYRAIPGVVKNKLEE 597

QY 81 IMLGVVIPDEEV-----GLGMELGYALSOGKYVLLVI-----PBEDYG 119
DB 598 QYNKIVIPNEENVVIYKIRQRLAKLGKEIEYIHKPKYCLPFLQPLQRLVKVKNEDDDFG 657

QY 120 KPINLMSWGSVDNVKMSQLK 140
DB 658 -----WGVVNVNFSKSNVK 671

RESULT 15
Q4ZQO6.PLABE PRELIMINARY; PRT; 264 AA.
AC Q4ZQO6;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Hypothetical protein (Fragment).
GN ORFNames=PB000574.01.0;
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5821;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karas M., Raine J.D., Carlton J.M., Kooij T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAAI01001273; CAH96134.1; -; Genomic DNA.
KW Hypothetical protein.
FT NON TER 264 264
SQ SEQUENCE 264 AA; 31399 MW; EDD7A47344723240 CRC64;

Query Match 11.0%; Score 93.5; DB 2; Length 264;
Best Local Similarity 27.1%; Pred. No. 8.3;
Matches 39; Conservative 24; Mismatches 52; Indels 29; Gaps 7;

QY 19 KAYKEAMEALKENPTIDLENSVYPLDNOYKIGIR-VDEHPEYLHDKVWATATYNNDLNGIK 77
DB 34 RSYLQAMECLKNNEM---NTYKKNKKNKYLDTKNLEDSDDLHKIKWINNNSINNKNSFK 90

QY 78 T-----NDIMLGVYIPDEEDVGLGMELGYALSOGKYVLLVIPDEYKPKINLMSWGSVDN 132
DB 91 TIQKYGNNI--GFLKDER-----RLNVALTRAKDSLWIIGDK-----TNLQKNSTWDS 136

QY 133 VIKMSQLK-----DFNFKPRFD 150
DB 137 LIKVAIARNCYVNLNLFDRSTKD 160

Search completed: December 23, 2005, 22:04:15
Job time : 238 secs

```